The Econometrics of Financial Comovement

Silde, E.

2017

document version
Publisher's PDF, also known as Version of record

The Econometrics of Financial Comovement

by

Erkki Silde

ISBN 978 90 5170 991 9

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. 691 in the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

VRIJE UNIVERSITEIT

The Econometrics of Financial Comovement

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Economische Wetenschappen en Bedrijfskunde
op vrijdag 2 juni 2017 om 9.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Erkki Silde

geboren te Tallinn, Estland

promotor:	prof.dr. A. Lucas
copromotoren:	dr. L.F. Hoogerheide
	dr. F. Blasques Albergaria Amaral

Acknowledgments

The birth of thesis can be blamed on several extraordinary circumstances. There is plenty to be grateful of and the below disclaimer serves to give partial credit where more than full credit is due.

First and foremost I am grateful that my laptop, an Intel core i5-3210M CPU that has survived the critical tasks to complete the dissertation project.

More humanistically, I would like to thank the committee members for agreeing to spend their time on my thesis: Prof. Dr. H.P. Boswijk, Prof. Dr. S.J. Koopman, Dr. A. Pick, Dr. A.H. Siegmann and Dr. C. Zhou. A special thanks goes to the (co-)promotors: Prof. Dr. A. Lucas, Dr. F. Blasques and Dr. L.F. Hoogerheide have provided invaluable feedback and support in the guidance and organization of the dissertation project. During the last couple of years, the influence of the committee members and (co-)promotors has been highly valuable.

Furthermore, the facilities and support of Tinbergen Institute and Duisenberg School of Finance staff have been excellent. Many of the faculty members and seminar speakers at Tinbergen Institute, VU University Amsterdam and Duisenberg School of

Finance have been patient and supportive throughout. In particular, I have benefitted from valuable insights of Svetlana Borovkova, Charles Bos, Peter Boswijk, Siem-Jan Koopman, Katarzyna Lasak. The dissertation project marked a phase during which many young econometricians were working in physical vicinity on similar topics. The potential of such an environment is truly appealing and the group of doctoral and postdoctoral students at VU has provided a valuable community to exchange ideas.

Life in the Netherlands has been adventurous, in the past years I have made plenty of new discoveries: meeting truly inspiring new people, hosting and visiting old friends from all over, trying out all sorts exotic foods, learning to appreciate German politeness, learning to appreciate both sun and rain. To top things off, I have come across a truly remarkable specimen called Lin.

The dissertation work heavily leans on previously given background work. First, I owe it all to the support of my family – they have had the patience to reason with me from an early age. My interest in conducting research probably stems from such a mindset. Second, I am indebted to the undergraduate studies at the University of Konstanz – not only providing a first class education in mathematics and quantitative economics as tools to make a living for the rest of my life, but also giving the sense of integrity and a moral compass to navigate through academic sidewinds. In particular, the opportunity to work for Prof. Dr. Dr. h.c. Günter Franke was a true insipration to pursue scientific research. The Tinbergen Institute graduate curriculum has also had its highlights – Siem-Jan Koopman and Herman van Dijk are worth a special

mention for introducing stimulating research-frontier curriculum into the Tinbergen Econometrics MPhil program. Young researcher support of the Society of Financial Econometrics and Journal of Applied Econometrics has been of great benefit both financially and ideologically.

Even if many of the questions that I have asked myself have proved too difficult to answer, there is still an upside of the limited success by having opened the door for further insights. This thesis organically documents such a journey in the context of some of my work on score driven models.

Erkki Silde

Tallinn, Estonia

August 2015

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Timely risk monitoring in the context of multivariate time series is central for allocation decisions in financial markets, yet the quest for appropriate models is still an ongoing one, both in terms of theory and empirics. The challenge of modeling multivariate time series can be compared to walking on a tightrope. The goal is to get to the other side, i.e. to uncover the rich dynamics by making use of the abundant data that has become available to us in the recent decade. We would expect that more data would immediately result in more information and a better understanding, yet we witness how the available models may become less reliable in such settings, both in terms of the computational burden and the statistical properties.

Particular concerns that often reappear in financial data are the nonlinear nature of payoffs and the time variation of market descriptions, say, due to the disruptive price behavior of markets during crisis times when compared to periods of

tranquil business conditions. An often employed approach that can capture multivariate dynamics in financial markets is the path of state space modeling (see also Durbin and Koopman, 2012). An alternative attempt that requires less computational sophistication while maintaining a systematic modeling framework is the class of score driven models that have been recently proposed to strike the balance between numerical feasibility and statistical efficiency. Within economics, the approach is known as the Generalized Autoregressive Score (GAS) framework of Creal, Koopman, and Lucas (2013), while signal processing literature refers to it as adaptive filtering (see also Goodwin and Sin, 2014). Section 1.1 presents the score driven context as a general framework and discusses the example of modeling dynamic correlations. Section 1.2 describes some of the similarities of state space models with the score driven framework. Section 1.3 collects the contributions of the thesis.

1.1 Score driven filtering framework

Consider the generic setting of data y_t being described by observation density $p(y_t; f_t, \theta)$, where f_t and θ denote the time-varying and static parameters, respectively.

Score driven models can be cast in the following form

$$y_t \sim p(y_t; f_t, \theta) \tag{1.1}$$

$$s_t(f_t; \lambda) = \mathcal{I}(f_t; \lambda)^{-a} \nabla(f_t; \lambda) \qquad a \in \{0, 1/2, 1\} \tag{1.2}$$

$$\nabla(f_t; \lambda) = \frac{\partial \log p(y_t; f_t, \theta)}{\partial f_t} \tag{1.3}$$

$$\mathcal{I}_t(f_t; \lambda) = \mathbb{E}_{t-1}\big[\nabla(f_t; \lambda)\nabla^\top(f_t; \lambda)\big] \tag{1.4}$$

$$f_{t+1} = \phi_y(y_t; f_t, \theta) = \phi_\nabla\left(s_t; f_t, \theta\right), \tag{1.5}$$

where $\mathbb{E}_t[\cdot]$ denotes the conditional expectation based on information available at time t and the recursion is initialized at a starting value f_1. The information matrix scaling power a has been customarily set to either 0, 1/2 or 1 for purposes of analytical or numerical convenience, though more research on the impact of the parameter choice would further clarify the theoretical underpinnings of the approach. The functional form $\phi_\nabla(\cdot)$ typically designates an intuitive functional form (e.g. of an ARMA-type such as the recursion discussed in Section 1.1.1 further below) to introduce information into the system via the scaled score s_t. Importantly, the GAS model provides one with a systematic functional form to update time-varying parameters f_t. Notably, the log-likelihood function

$$\mathcal{L}_T = \sum_{t=1}^{T} \log p\big(y_t; f_t, f_1, \theta\big)$$

depends on the hyper-parameter θ, where the updating recursion of the time-varying parameter f_t is a measurable function of the past scaled scores $s_t(f_t; \lambda)$ and therefore of y_t. This classifies the specification in (1.1)–(1.5) as an 'observation driven model' in the terminology of Cox (1981). In other words, having conditioned on the data and knowing the parameters θ, one is able to exactly reconstruct the time-varying parameter path and the likelihood, which constitutes the main computational appeal of the model.

1.1.1 Modeling dynamic correlations

Consider the example of a portfolio allocation problem. The key to such a financial decision making situation lies in the risks associated with alternative choices. In turn, if one's portfolio consists of multiple assets, then their correlations determine the degree of diversification of the overall exposure. However, as correlations are known to be subject to changes over time (Engle, 2002a), risk monitoring hinges upon accurate correlation estimation procedures.

The leading motivating example is the bivariate dependence model for financial

returns, denoted by y_t,

$$\Sigma_t = \mathbb{V}\text{ar}[y_t] \tag{1.6}$$

$$\Sigma_t = \begin{pmatrix} \sigma_{1,t}^2 & \rho(f_t)\sigma_{1,t}\sigma_{2,t} \\ \rho(f_t)\sigma_{1,t}\sigma_{2,t} & \sigma_{2,t}^2 \end{pmatrix} \tag{1.7}$$

Fortunately, there is extensive research available for determining the volatility parameters $\sigma_{1,t}$ and $\sigma_{2,t}$ (Andersen et al., 2007; Barndorff-Nielsen, 2002; Corsi, 2009; Hansen et al., 2012; Harvey, 2013). We focus on different GARCH-type models to explore the dynamics of the time-varying correlation parameter $\rho(f_t)$, where $\rho : \mathbb{R} \to [-1, 1]$ is a transformation of the time varying parameter f_t.

We next apply the score driven framework to obtain an updating recursion for the dynamic correlation parameter of a Student's $t(\nu)$ distribution. We further parametrize the correlation by $\rho_t = \rho(f_t) = \tanh(f_t)$ and do not scale score by the information matrix.[1] Following the definitions of the score driven framework outlined in the previous Section 1.1, we can now derive the resulting updating equation for

[1]The information matrix choice is further motivated in Chapter 3.

the dynamic correlation parameter. This leads to the following system of equations

$$f_{t+1} = \omega + \beta f_t + \alpha s(\rho_t, y_t; \lambda), \qquad f_1 = \omega/(1-\beta) \tag{1.8}$$

$$s(\rho_t, y_t; \lambda) = \frac{1}{2}\frac{1}{1-\rho_t^2} \cdot \Big[w_t(1+\rho_t^2)(y_{1t}+y_{2t})^2 - w_t(1+\rho_t)^2(y_{1t}^2+y_{2t}^2) + 2\rho_t(1-\rho_t^2) \Big] \tag{1.9}$$

$$w_t = \frac{\nu+2}{\nu-2+y_t^\top \Sigma_t^{-1} y_t},$$

see also chapters 2 and 3 for a more detailed discussion of this model.

It has been convincingly argued by Creal et al. (2013) that due to the fat tailed nature of financial returns as modeled by a Student's t(ν) distribution, the functional form of $s(\rho_t, y_t; \lambda)$ should exhibit a robustness feature of bounded news impact curve. This contrasts with a Gaussian distribution for which outliers can have an unbounded influence on inference regarding the correlation parameter. Such a robustness feature has been argued to be crucial due to the limited capability of our models to describe economic forces and the ever so occasional 'sigma-events' (Calvet et al., 2014), i.e. events that appear unexpected given the manner that real-life events may have been modeled. For instance, the buildup of financial crises or sudden moves on currency markets may be difficult predict, while the observed data exhibits extreme movements, therefore creating a substantial gap between reality and the corresponding model world.

In order to provide a motivating illustration of this point, we consider alterna-

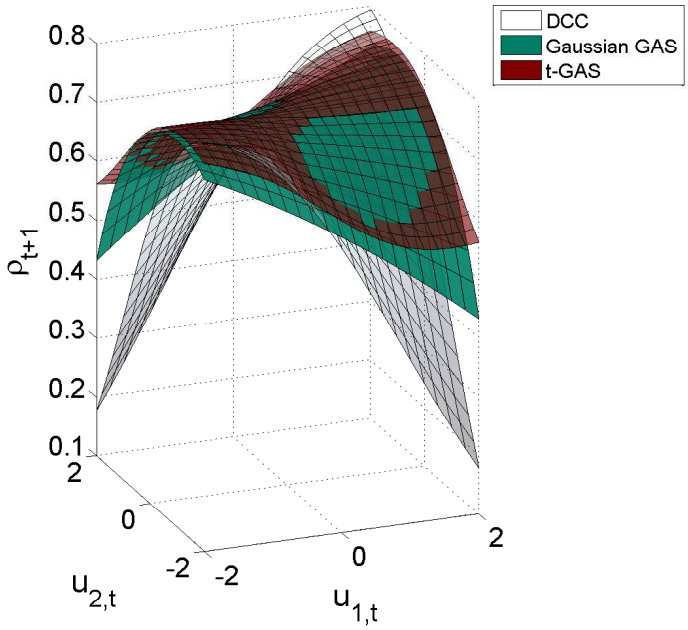


Figure 1.1: News impact surface of persistent, mean-reverting bivariate correlation models

Notes: The horizontal axes depict realizations of i.i.d. noise, while the vertical axis marks filtered correlation prediction for the next period.

tive correlation updating mechanics. We explore three models: the (partly linear) Dynamic Conditional Correlation (DCC) model, a Gaussian GAS copula model and finally, a dynamic GAS copula model with fat tails. Figure 1.1 displays the correlation 'news impact surface' of a bivariate model (1.8) – (1.9), i.e., the impact of two normalized returns on their updated contemporaneous Pearson correlation coefficient. It becomes apparent that the heavy-tailed GAS specification is more resilient than the DCC specification in case of a single, out-of-the-model event.

1.2 Intuitive properties of GAS models

Complex economic data generating processes require tools that would adequately describe them in an attempt to unravel the underlying mechanics. In fact, White (see Granger, 2001, p.470) argues as follows:

> One clear conclusion that can be reached is that one can never be sure that a white noise is not forecastable, either from some nonlinear or time-varying model or from the use of a wider information set, so one should never stop trying to find superior models.

According to White's theorem (see also Granger, 2008), any nonlinear model can be approximated by a time-varying parameter linear model. The GAS framework provides an attractive tool to study economic phenomena due to the appeal of a computationally simple way to a) make parameters time varying; b) accommodate non-linear dynamic relationships. In this way, GAS models provide a relatively straightforward way to build approximating models in the sense of White.

However, while empirically and intuitively compelling, theoretical underpinnings of GAS models have remained less understood prior to the start of the thesis project. To gain a deeper understanding of the issues, we compare the behavior of GAS models with computationally more involved non-linear non-Gaussian state space models, also known as parameter driven models as the time-varying parameters constitute independent processes and are not necessarily functions of observed data.

Consider the example of a linear dynamic location model with measurement error.

$$y_t = f_t + \sigma_u u_t \tag{1.10}$$

$$f_t = \omega + \beta f_{t-1} + \alpha v_t, \qquad (u_t, v_t) \sim \mathrm{N}\left(0_2, \mathrm{I}_2\right) \quad \text{i.i.d.} \quad . \tag{1.11}$$

We can impose the stationarity constraint $|\beta| < 1$ and compute the moments of y_t and f_t analytically, see also Table 1.1. Now consider the GAS counterpart to the model (1.10) – (1.11) as defined by

$$\begin{aligned} y_t &= f_t + \sigma_u u_t & u_t &\sim \mathrm{N}\left(0,1\right) \quad \text{i.i.d} \\ f_{t+1} &= \omega^\dagger + \beta^\dagger f_t + \alpha^\dagger s_t(f_t;\lambda) & s_t &= \mathcal{I}_t(f_t;\lambda)^{-\frac{1}{2}} \nabla(f_t;\lambda) \\ \nabla(f_t;\lambda) &= \frac{y_t - f_t}{\sigma_u^2} & \mathcal{I}_t(f_t;\lambda) &= \frac{1}{\sigma_u^2}. \end{aligned}$$

where we have used the square root information matrix scaling ($a = 1/2$) to ensure that the innovations to the state transition have unit variance.[2]

The moments for fixed $\theta = (\omega, \alpha, \beta)$ and $\theta^\dagger = (\omega^\dagger, \alpha^\dagger, \beta^\dagger)$, are reported in Table 1.1. The striking pattern is that the moments of the data can be computed by identical functions that map θ and $\theta^\dagger$ into their respective population moments. For the sake of exposition, assume that σ_u^2 is known. Then, in the just-identified case,

[2]The square root scaling can also be justified by the Jeffreys' principle, which in this context requires that the updating rule is invariant under reparametrizations of the time-varying parameter. The model with square root information matrix scaling also carries a special role in this thesis: in Chapter 3 the formulation enforces applicability of the Markov Chain theory, while in Chapter 4 it is the natural starting point of considering time variation in model mis-specification.

	Linear State Space Model (1.10)–(1.11)	GAS counterpart with $a = 1/2$
$\mathbb{E}y_t$	$\mathbb{E}f_t$	$\mathbb{E}f_t$
$\mathbb{V}\mathrm{ar}\left[y_t\right]$	$\mathbb{V}\mathrm{ar}\left[f_t\right] + \sigma_u^2$	$\mathbb{V}\mathrm{ar}\left[f_t\right] + \sigma_u^2$
$\mathbb{E}y_t y_{t-1}$	$\beta\mathbb{E}\left[f_t^2\right] + \omega\mathbb{E}\left[f_t\right]$	$\beta^\dagger\mathbb{E}\left[f_t^2\right] + \omega^\dagger\mathbb{E}\left[f_t\right]$
$\mathbb{E}f_t$	$\omega/(1-\beta)$	$\omega^\dagger/(1-\beta^\dagger)$
$\mathbb{V}\mathrm{ar}\left[f_t\right]$	$\alpha^2/\left(1-\beta^2\right)$	$\left(\alpha^\dagger\right)^2/\left(1-\left(\beta^\dagger\right)^2\right)$

Table 1.1: Comparison of the unconditional moments implied the linear models

	Stochastic Volatility Model (1.12)–(1.13)	GAS counterpart with $a = 1/2$
$\mathbb{E}y_t$	0	0
$\mathbb{V}\mathrm{ar}\left[y_t\right]$	$\mathbb{E}\exp(f_t)$	$\mathbb{E}\exp(f_t)$
	$=\exp(\mathbb{E}\left[f_t\right] - \frac{1}{2}\mathbb{V}\mathrm{ar}\left[f_t\right])$	$=?$
$\mathbb{E}\log y_t^2$	$\mathbb{E}f_t + \mathbb{E}\log u_t^2$	$\mathbb{E}f_t + \mathbb{E}\log u_t^2$
$\mathbb{V}\mathrm{ar}\log y_t^2$	$\mathbb{V}\mathrm{ar}\left[f_t\right] + \mathbb{V}\mathrm{ar}\left[\log u_t^2\right]$	$\mathbb{V}\mathrm{ar}\left[f_t\right] + \mathbb{V}\mathrm{ar}\left[\log u_t^2\right]$
$\mathbb{C}\mathrm{ov}\left(\log y_t^2, \log y_{t-1}^2\right)$	$\beta\mathbb{V}\mathrm{ar}\left[f_t\right]$	$\beta^\dagger\mathbb{V}\mathrm{ar}\left[f_t\right]$
$\mathbb{E}f_t$	$\omega/(1-\beta)$	$\omega^\dagger/(1-\beta^\dagger)$
$\mathbb{V}\mathrm{ar}\left[f_t\right]$	$\alpha^2/\left(1-\beta^2\right)$	$\left(\alpha^\dagger\right)^2/\left(1-\left(\beta^\dagger\right)^2\right)$

Table 1.2: Comparison of the unconditional moments implied by non-linear models

equating three of the sample moments is sufficient for consistent estimates.

Importantly, the moments of nonlinear and possibly non-Gaussian state space models also find their counterpart in the observation driven framework. Consider the plain vanilla stochastic volatility model

$$y_t = \exp\left(\tfrac{1}{2}f_t\right)u_t \tag{1.12}$$

$$f_t = \omega + \beta f_{t-1} + \alpha v_t, \qquad (u_t, v_t) \sim \mathrm{N}\left(0_2, \mathrm{I}_2\right) \quad \text{i.i.d.} \quad . \tag{1.13}$$

In the square root information scaling GAS counterpart of this model, the scaled score is defined by $s_t = \left(\exp(-f_t)y_t^2 - 1\right)/\sqrt{2}$. We notice in Table 1.2 that also in this model, some moments of the data can be calculated by the same function of the parameters of either the State Space model or its GAS counterpart.

However, the fact that the *simulated* moments are similar for both models, does not necessarily imply that both models would behave similarly when *filtering* the observed data with a particular potentially mis-specified model. Consider the example of observing data being generated by the parameter driven model (1.12)–(1.13) which we denote by $\mathcal{P}_0$, while applying the GAS model to filter the observed data, denoted by $\mathcal{P}^\dagger$. In this example, the GAS filtering recursion of the log-variance $f_t^\dagger = f_t^\dagger(y_{t-1}; \theta^\dagger)$, is given by

$$f_{t+1}^\dagger = \omega^\dagger + \beta^\dagger f_t^\dagger + \alpha^\dagger \frac{1}{\sqrt{2}} \left(\frac{\exp f_t}{\exp f_t^\dagger} u_t^2 - 1 \right).$$

where $\theta^\dagger$ is determined by minimizing the expected Kullback-Leibler divergence between the probability distribution implied by parameter driven model $\mathcal{P}_0$ and the probability distribution implied by its GAS counterpart $\mathcal{P}^\dagger$, which, in our volatility modeling example can be expressed as

$$D_{\mathrm{KL}}(\mathcal{P}_0 \| \mathcal{P}^\dagger) = \tfrac{1}{2}\mathbb{E}\left[d_t(\theta, \theta^\dagger) + \exp\left(-d_t(\theta, \theta^\dagger)\right) - 1 \right]$$

$$d_t(\theta, \theta^\dagger) = f_t^\dagger(\theta^\dagger) - f_t(\theta)$$

As intuitive as the above recursions may seem, they are classified as nonlinear stochastic difference equations which may be difficult to grasp analytically and for a large part of this thesis, require a numerical treatment. Some questions immediately arise: Do recursions such as the ones for $f_{t+1}^{\dagger}$ and $d_t(\theta, \theta^{\dagger})$ converge in distribution? What are the meaningful restrictions on the parameters $\theta^{\dagger}$? How to recognize potential mis-specification of the updating recursion to adapt the functional form of ϕ_{∇}? Is there a sense in which certain observation driven models are designed to have the minimal expected Kullback-Leibler divergence $D_{\mathrm{KL}}(\mathcal{P}_0 \| \mathcal{P}^{\dagger})$ to parameter driven alternatives? These are some of the common issues that arise when we analyze novel observation drive models. The short theoretical background described above is useful to keep in mind when putting the methodological and empirical contributions of the thesis into a perspective: we are comparing descriptions of the data with the corresponding properties of the model.

This is where the motivation for this thesis ends and the real work starts. The following chapters aim to address such questions or hopefully, can provide results that would help answer these questions some day. We now describe the main contributions of each of the chapters.

1.3 Outline of thesis

This thesis consists of four main chapters. Each chapter discusses different aspects of multivariate score driven filtering models. While each chapter is sufficiently self-contained, there is some benefit in reading the chapters sequentially as the exposition and notation are fairly linear throughout the thesis.

Chapter 2 (based on joint work with F. Blasques and A. Lucas) extends the prevailing approach of univariate stochastic recurrence equations to a GAS correlation model. We are interested in verifying circumstances under which the model is stationary such that we would be able to conduct standard asymptotic inference. Yet conditions arising from stochastic fixed point theorems are analytically impossible and numerically difficult to handle, so we resort to worst case contraction conditions. Complications arise due to multivariate matrix decompositions being not uniquely defined and this chapter establishes a solution to this indeterminacy in the stochastic recurrence equations framework. Furthermore, we find that this indeterminacy can decide whether the theoretical assumptions are empirically applicable.

Chapter 3 proposes an alternative approach to describe the stability of multivariate systems for a class of observation driven models, which also entails GAS correlation specifications. Theoretically, we establish an estimation routine that is robust to misspecification of the observation density. Empirically, we establish a complete model to extract information from option markets to help us pin down correlation levels more precisely and to infer the time-varying correlation risk premia.

Chapter 4 addresses the important question of adjusting score driven updating recursions to account for static and time-varying sources of mis-specification. As an example for static misspecification, we consider the leverage effect in the score driven volatility news impact curve. In other words, we incorporate the observation that negative equity returns are associated with higher future volatility than positive returns into the score driven filtering framework. As an example for time-varying sources of misspecification, we consider models with time-varying long run correlation levels.

Chapter 5 touches upon a further central theme in modeling multivariate dynamic systems: striking a balance between parsimony and specification flexibility. The former is needed in order to avoid the abyss of unreliable inference about the underlying dynamics when the model can become "too big" relative to the information content of the data. The latter is needed not to miss any of the central driving mechanisms in the data by making overly simplified or restrictive assumptions. Whereas in Chapter 3 we stressed parsimony and studied a highly parsimonious and sufficiently interpretable model with a minimum number of time-varying parameters, in Chapter 5 I present a much more flexible model for multivariate time series dynamics. Despite its flexibility, the model is still highly parsimonious in its dynamics and concisely captures the main features of the data. I apply the model to investigate equity market contagion, extending the contagion analysis of Diebold and Yilmaz (2009) to a setting with much more dynamic flexibility of the underlying time series model, and

thus much more flexibility of for the underlying contagion patterns.

Chapter 6 summarizes the contributions in a more detailed manner and puts the results in a wider perspective.

Chapter 2

A Stochastic Recurrence Equations Approach to Score Driven Dynamic Correlation Models

Abstract. We describe worst-case contraction conditions for a recently proposed class of score driven dynamic correlation models. These models have important applications in empirical work. The contraction regions are derived by restricting the conditions in Bougerol (1993) for analytic and computational convenience. The worst-case contraction regions take a non-standard form and in particular imply strict stationarity and ergodicity (SE). We show that the non-standard shape of the sufficiency regions cannot be avoided by reparameterizing the model or by rescaling the score steps in the transition equation for the correlation parameter. This makes the result markedly different from the volatility case. Observationally equivalent decompositions of the stochastic recurrence equation yield regions with different shapes and sizes. We use these results to establish the consistency and asymptotic normality of the maximum likelihood estimator. We illustrate our results with an analysis of time-varying correlations between UK and Greek equity indices. We find that also in empirical applications different decompositions can give rise to different conclusions regarding the stability of the estimated model.

2.1 Introduction

Time-variation in correlations is an important feature of economic and financial data and a crucial ingredient of empirical analyses, such as the assessment of risk and the construction of investment portfolios. Available models for capturing the time-variation in correlations include, amongst many others, the BEKK model of Engle and Kroner (1995b), the switching correlation models of Pelletier (2006), the DCC model of Engle (2002b) with its adaptation by Aielli (2013), the DECO model of Engle and Kelly (2012), the dynamic copula models of Patton (2009) and Oh and Patton (2012), and the score driven models of Creal et al. (2011, 2013) and Harvey (2013); see also the overviews of Silvennoinen and Teräsvirta (2009) and Bauwens et al. (2006).

We focus on the stochastic properties of the recently proposed score driven models of Creal et al. (2011, 2013) and Harvey (2013), which we refer to as generalized autoregressive score (GAS) models. These models have proved particularly useful when modeling fat-tailed or skewed data, such as often encountered in empirical finance; see for example Janus et al. (2014), Harvey and Luati (2014), and Lucas et al. (2014). The dynamics of correlations and volatilities in these models are driven by the score of the error distribution. If the latter is fat-tailed, the score driven dynamics automatically correct for influential observations, see Creal et al. (2011). In this

This chapter is largely based on Blasques et al. (2016).

way, they share some similarities with models from the robust GARCH literature; see for example Boudt et al. (2013). The score driven approach used in the construction of GAS models, however, provides a much more general and unified framework for parameter dynamics that is applicable far beyond the volatility and correlation context; see Creal et al. (2013, 2014) for a range of other examples. In addition, from a forecasting perspective GAS models often have a similar performance to correctly specified state-space models, see Koopman et al. (2012).

Despite their proven empirical usefulness, the theoretical properties of GAS models are less well developed. The complication lies in the highly nonlinear transition dynamics of the time-varying parameter in typical GAS models. In this paper we contribute to our understanding of the stochastic properties of GAS models for dynamic correlations. The fundamental question is to understand which parameterizations, and parameter values generate stationary and ergodic (abbreviated as SE from now on) time series processes. This offers an important characterization of the stochastic properties of GAS models. SE properties can be combined with the existence of unconditional moments for the objective function to establish proofs of consistency and asymptotic normality of extremum estimators; see e.g. Straumann and Mikosch (2006) for maximum likelihood estimation of nonlinear conditional volatility models, Francq and Zakoian (2011) and Boussama et al. (2011) for the case of multivariate GARCH models, and Harvey (2013) for GAS volatility models. For each correlation model we consider, we identify the parameter values that ensure the SE property and

call this the 'SE region' of the parameter space. To establish SE regions, we follow the classical average contraction argument for stochastic recurrence relations as laid out in the sufficient conditions formulated by Bougerol (1993). Given these conditions, we compute numerically the SE regions for a range of empirically relevant models.

We have four contributions. First, we are the first to derive SE regions for the class of score driven correlation models that have been suggested recently in the literature. We show that the SE sufficiency regions take a highly non-standard form, dissimilar to the familiar triangle and curved triangular shapes for the GARCH model; see Nelson (1990). In an empirical example, we demonstrate that the conditions for nonlinear recurrence equations can be used to ensure stationarity of concrete models, applied on real data. This also extends the results in Blasques et al. (2014b) for volatility and tail index models with univariate observations to the case of time-varying parameters and multivariate observations.

Second, we show that the shape and size of the SE sufficiency region as derived from the conditions of Bougerol (1993) depends on the way the stochastic recurrence equation for the correlation is constructed from bivariate uncorrelated noise. In particular, we show that the choice of the square root of the correlation matrix in this construction has a non-trivial effect on the size of the SE sufficiency region.

Third, we show analytically why the correlation case is markedly different from the volatility case. For the volatility case, Harvey (2013) shows that modeling the log-volatility renders the information matrix independent of the time-varying volatility.

The resulting stochastic recurrence equation becomes linear, and we can use linear process theory to study the SE properties. A similar feature is generally not available for the dynamic correlation model: neither a reparametrization of the correlation, nor a scaling of the score steps, makes the stochastic recurrence equation a linear process. The reason is that unlike the volatility case, the GAS steps for the correlation model consist of two separate terms with different nonlinearities in the correlation parameter.

Fourth, we use our SE results to establish the consistency and asymptotic normality of the ML estimator.

The remainder of this paper is organized as follows. In Section 3.2, we introduce our model for dynamic bivariate correlations. In Section 2.3 we state the conditions for the SE sufficiency regions. In Section 2.4 we establish model invertibility as well as the consistency and asymptotic normality of the ML estimator. In Section 3.4, we determine the SE regions numerically for a number of different models and provide an empirical illustration for UK and Greek equity indices. We conclude in Section 3.5. The appendix gathers the more technical results and derivations.

2.2 Score driven models for correlations

Consider a real-valued bivariate stochastic sequence of observations $\{y_t\}_{t\in\mathbb{N}}$ generated by a zero mean elliptical conditional distribution with a density function of

the form

$$\frac{p(y_t^\top R(f_t)^{-1} y_t)}{|R(f_t)|^{1/2}}. \tag{2.1}$$

Here $p : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ denotes the so-called real-valued density generator function in the quadratic form $y_t^\top R(f_t)^{-1} y_t$. For example, if y_t is conditionally normal, we have $p(x) = (2\pi)^{-1} \exp(-x/2)$. We fully focus the exposition on the correlation case by restricting the variances in $R(f_t)$ to unity.[1] Note that due to the ellipticity, time-varying conditional distributions are entirely described by the correlation matrix

$$R(f_t) = \begin{pmatrix} 1 & \rho(f_t) \\ \rho(f_t) & 1 \end{pmatrix},$$

where $\rho(f_t) \in [-\delta, \delta]$ with $\delta \in (0, 1)$ is the dynamic correlation parameter at time t and $\{f_t\}_{t \in \mathbb{N}}$ is real-valued sequence for the time-varying parameter.

The formulation in (2.1) can also be interpreted as a copula model, see the discussion in Patton (2009). Under the assumptions of stationary marginals and no volatility spillovers, stability conditions for the copula then lead to stability of the whole model. The class of elliptical models is also economically interesting, as it enables an analytic characterization of the resulting portfolio returns and the risk-return tradeoff; see for example Chamberlain (1983), Owen and Rabinovitch (1983), and Hamada and Valdez (2008).

Following Creal et al. (2011, 2013), the generalized autoregressive score (GAS)

[1] Time-varying variances in score driven models have already been dealt with in for example Creal et al. (2011) and Harvey (2013).

dynamics for the time-varying parameter f_t in (2.1) take the form

$$f_{t+1} = \omega + \beta f_t + \alpha s_t(f_t, y_t), \qquad t = 1, 2, \ldots, \tag{2.2}$$

$$s_t(f_t, y_t) = S(f_t) q(y_t, f_t), \qquad q(y_t, f_t) = \frac{\partial}{\partial f} \log \frac{p(y_t^\top R(f_t;\lambda)^{-1} y_t)}{|R(f_t;\lambda)|^{1/2}}\bigg|_{f=f_t}, \tag{2.3}$$

with an arbitrary fixed initial condition $f_1 \in \mathcal{F}$. We define the parameter vector $\theta \in \Theta$ as $\theta = (\omega, \alpha, \beta, \lambda)$, where $(\omega, \alpha, \beta) \in \mathbb{R}^3$ is a vector of updating parameters, and $\lambda \in \mathbb{R}^{n_\lambda}$ allows $n_\lambda \geq 0$ density tail shape parameters to be estimated. The time-invariant parameter space is described by $\Theta \subseteq \mathbb{R}^{3+n_\lambda}$. We suppress dependence of the scaling and the score on θ by writing $s_t(f_t, y_t) \equiv s_t(f_t, y_t; \lambda)$, $S(f_t) \equiv S(f_t; \lambda)$, and $q(y_t, f_t) \equiv q(y_t, f_t; \lambda)$. For the case of the bivariate correlation model (2.1), we obtain

$$q(y_t, f_t) = \frac{\dot{\rho}(f_t)}{1-\rho(f_t)^2} \left(\dot{p}\left(y_t^\top R(f_t)^{-1} y_t\right) \left(2\rho(f_t) y_t^\top R(f_t)^{-1} y_t - y_t^\top \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} y_t \right) + \rho(f_t) \right) \tag{2.4}$$

with $\dot{p}(x) = \partial \log p(x)/\partial x$ and $\dot{\rho}(f_t) = \partial \rho(f_t;\lambda)/\partial f|_{f=f_t}$.

Each choice for the scaling function S in (2.3) gives rise to a new GAS model. An often used choice of S relates to the local curvature of the score as measured by the information matrix, for example

$$S(f_t;\lambda) = (\mathcal{I}_t(f_t;\lambda))^{-a}, \qquad \mathcal{I}_t(f_t;\lambda) = \mathbb{E}_{t-1}[q(y_t, f) q(y_t, f)^\top], \tag{2.5}$$

where a is typically taken as 0, 1/2 or 1.

The parameter dynamics in (2.2) and (2.3) are intuitive. The time-varying parameter f_t is updated in the (scaled) direction of steepest ascent as measured by the scaled conditional log observation density at time t. For example, standard GARCH and BEKK models are special cases of the GAS framework under normality, see Creal et al. (2013) The GAS set-up is very general and can also easily be applied outside the correlation context as long as a conditional observation density is available. For other examples, including many new models, we again refer to Creal et al. (2013, 2014).

2.3 Contraction conditions

We follow the approach of Blasques et al. (2014b), who consider a treatment of univariate GAS models. Our stationarity and ergodicity (SE) results build on the stochastic recurrence relations or iterated random functions approach; see Diaconis and Freedman (1999) and Wu and Shao (2004). In particular, we use the sufficient conditions of Bougerol (1993) and results in Straumann and Mikosch (2006) to establish, for any fixed initial condition $f_1 \in \mathcal{F}$, exponentially fast almost sure convergence of the time series $\{y_t, f_t(\theta, f_1)\}_{t\in\mathbb{N}}$ generated by (2.1)–(2.3) to a unique SE solution $\{y_t, f_t(\theta)\}_{t\in\mathbb{Z}}$.

Let $\mathcal{F} \subseteq \mathbb{R}^k$ and $\mathcal{Y} \subseteq \mathbb{R}^n$ denote the sets where f_t and y_t take values, respectively.

We have that $\rho : \mathcal{F} \to (-\delta, \delta)$ for $0 < \delta \leq 1$ and $s_t : \mathcal{F} \times \mathcal{Y} \times \Theta \to \mathbb{R}$ is almost surely (a.s.) smooth in all its arguments and Lipschitz in $f \in \mathcal{F}$. Interestingly, it turns out from the proof of Lemma 2 below that the Fisher (tanh) transformation is crucial in ensuring the Lipschitz property of the stochastic recurrence equation for the time-varying correlation parameter. Using the model as specified in (2.1)–(2.4), we analyze the stochastic properties of y_t and f_t via the stochastic recurrence equation

$$\begin{aligned} y_t &= h(f_t)u_t, \qquad h(f_t)h(f_t)^\top = R(f_t) \\ f_{t+1} &= \phi_t(f_t;\theta) := \phi\left(h(f_t)u_t, f_t;\theta\right) = \omega + \beta f_t + \alpha S(f_t)\, q(h(f_t)u_t, f_t) \qquad t \in \mathbb{Z}, \end{aligned} \tag{2.6}$$

and $\{u_t\} := \{u_t\}_{t\in\mathcal{Z}}$ is an independent and identically distributed (i.i.d.) sequence with $y_t = h(f_t)u_t$. We notice that the dynamics of $\{f_t\}$ in (2.6) are now written in terms of the innovation sequence $\{u_t\}$ rather than the observed data $\{y_t\}$ by substituting $h(f_t)u_t$ for y_t. As a result, when seen as a function of f, the shape of $q(h(f_t;\lambda)u_t, f)$, for every u_t, is markedly different from that of $q(y_t, f)$, for every y_t. This additional dependence on f may either complicate or simplify the nonlinear dependence of f_{t+1} on f_t as embedded in (2.6). Second, the functional form of (2.6) is not uniquely defined. Each square root $h(f_t;\lambda)$ of the correlation matrix $R(f_t;\lambda)$ leads to an observationally equivalent model in y_t. The choice of $h(f_t;\lambda)$, however, is not innocuous for determining the size and shape of the SE region, as we see later.

Continuity of ϕ_t in u_t for every t can be used to ensure that $\{\phi_t\}$ is an i.i.d. sequence

of functions. Together with equation (2.6), it then follows directly from Bougerol (1993) and Straumann and Mikosch (2006) that there is a unique SE solution to (2.1)–(2.3) if ϕ_t is contracting on average, i.e., if the Lyapunov exponent of the mapping is negative. In particular, we obtain the desired SE result if

$$\mathbb{E}\log \sup_{f,f^*\in\mathcal{F}} \frac{|\phi_t(f;\theta)-\phi_t(f^*;\theta)|}{|f-f^*|} \leq \mathbb{E}\log \sup_{f\in\mathcal{F}} \left|\frac{\partial \phi_t(f;\theta)}{\partial f}\right| \;<\; 0; \tag{2.7}$$

see Bougerol (1993). In computing the supremum in condition (2.7), f is treated as a parameter rather than as the random variable f_t.

For the score driven dynamic correlation model of Section 3.2, we prove the following result in the Appendix.

Lemma 1. *Let Ψ be a class of functions such that for every $\psi \in \Psi$, $\psi \in \mathcal{C}^1([-\delta,\delta],\mathbb{R})$ with $\dot{\psi}(\rho) = \partial\psi(\rho)/\partial\rho = \mathcal{O}((1-\rho^2)^{-1/2})$. Assume that $\mathbb{E}|\dot{p}(u_t^\top u_t)u_{i,t}u_{j,t}| < \infty$ for $i,j \in \{1,2\}$, with $u_t = (u_{1,t},u_{2,t})^\top$. For any fixed initial condition $f_1 \in \mathcal{F}$, the process $\{f_t(\theta,f_1)\}_{t\in\mathbb{N}}$ generated by the dynamic correlation model (2.1)–(2.4) converges exponentially fast almost surely[2] (e.a.s.) to a unique stationary and ergodic solution $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ if*

$$\inf_{\psi\in\Psi} \mathbb{E}\log \sup_{f\in\mathcal{F}} \Big| \beta + \alpha\left(\frac{\partial}{\partial f}\left(\frac{S(f)\,\dot{\rho}(f)}{1-\rho(f)^2}\right)\right) g(\rho)(f)k(u_t) + \alpha\frac{S(f)\,\dot{\rho}(f)^2}{1-\rho(f)^2}\dot{g}(\rho)(f)k(u_t)\Big| \;<\; 0, \tag{2.8}$$

[2] A sequence $\{x_t\}$ converges exponentially fast almost surely to a sequence $\{\tilde{x}_t\}$ if for some constant $c > 1$ we have $c^t \cdot |x_t - \tilde{x}_t| \overset{a.s.}{\to} 0$ for $t \to \infty$.

where

$$g(\rho) = \left(\rho \ , \ \rho c_{2\psi}(\rho) - \sqrt{1-\rho^2} s_{2\psi}(\rho) \ , \ \sqrt{1-\rho^2} c_{2\psi}(\rho) + \rho s_{2\psi}(\rho)\right), \quad (2.9)$$

$$k(u_t) = \left(\dot{p}(u_t^\top u_t) u_t^\top u_t + 1 \ , \ \dot{p}(u_t^\top u_t)(u_{1,t}^2 - u_{2,t}^2) \ , \ -2\dot{p}(u_t^\top u_t) u_{1,t} u_{2,t}\right)^\top, \quad (2.10)$$

$\dot{g}(\rho) = \partial g(\rho)/\rho$, $c_{2\psi}(\rho) = \cos(2\psi(\rho))$, *and* $s_{2\psi}(\rho) = \sin(2\psi(\rho))$.

We note several features of the result stated in Lemma 1. First, the SE region only depends directly on the parameters α and β, on the functional forms of $S(f_t; \lambda)$ and $q(h(f_t; \lambda)u_t, f)$, and on the density of u_t. The dependence on the latter enters in two ways, namely through the expectations operator in (2.8) and through the functional form of $k(u_t)$ in (2.10). Also note that the expectations operator in (2.8) does not necessarily require the second moments of u_t to exist. Instead, we only require the expectation of $|\dot{p}(u_t^\top u_t) u_{i,t} u_{j,t}|$ for $i, j \in \{1, 2\}$ to exist. This condition is much weaker, particularly for fat-tailed elliptical densities. For example, it is easily satisfied for the bivariate Cauchy distribution, even though neither the second, nor the first moment exists for this distribution. The continuity and boundedness properties of s_t can be verified immediately for parametric distributional forms, notably for the Student's t density in Section 2.5.1.[3] Therefore, condition (2.8) effectively forms a sufficient condition for the SE property of the model.

Second, equation (2.8) directly reveals that the correlation case is markedly differ-

[3]The functional forms for the updating equation for the particular case of the Student's t distribution are presented in Appendix 2.B.

ent from the volatility case. For the volatility case, it is shown in Harvey (2013) and Blasques et al. (2014b) that through a clever choice of parameterization h or scaling S the scaled score in recurrence relation (2.6) can be made independent of f_t. The SE condition then reduces to the requirement that $|\beta| < 1$. In the volatility case the analogue of the function $g(\rho)$ in (2.9) is scalar valued. In the correlation case, equation (2.8) shows that through the trivariate nature of the functions $g(\rho)$ and $\dot{g}(\rho)$ the contraction condition consists of a number of different terms, each with a different nonlinear dependence on f. It is impossible to off-set all of these simultaneously by a single choice of scaling function or parameterization. This makes the correlation model theoretically more interesting in its own right.

Third, the SE sufficient condition in (2.8) has an additional degree of flexibility provided by the choice of ψ. As follows from the proof of Lemma 1, the function ψ determines which square root $h(f_t;\lambda)$ is used for the correlation matrix $R(f_t;\lambda)$. For the purpose of guaranteeing a proper correlation matrix, define $\xi(\rho) = \arcsin(\rho) - \psi(\rho)$, and

$$h(f_t;\lambda) = \begin{pmatrix} \cos(\xi(\rho(f_t;\lambda))) & \sin(\xi(\rho(f_t;\lambda))) \\ \sin(\psi(\rho(f_t;\lambda))) & \cos(\psi(\rho(f_t;\lambda))) \end{pmatrix}, \qquad \psi(\rho(f_t;\lambda)) = k_\psi \cdot \arcsin(\rho(f_t;\lambda)), \tag{2.11}$$

for some constant $k_\psi \in \mathbb{R}$, such that $h(f_t;\lambda)h(f_t;\lambda)^\top = R(f_t;\lambda)$ for all $\psi \in \Psi$. This parametrization gives rise to familiar alternatives for matrix roots: For $k_\psi =$

1/2 we obtain the transposition-symmetric matrix root, whereas the choice $k_\psi = 1$ reduces to the Cholesky decomposition with $y_{1,t} = u_{1,t}$ and $y_{2,t} = \rho_t u_{1,t} + \sqrt{1-\rho_t^2}u_{2,t}$. Any choice of ψ and thus of h results in an observationally equivalent model for y_t. The dynamic properties of $\{f_t\}$ following from (2.8), however, depend on the precise ψ in the functional form of $\phi(\cdot)$ that is chosen. We therefore obtain a sufficient condition for SE if (2.7) is satisfied for some choice of $\psi \in \Psi$ satisfying the conditions formulated in Lemma 1. This yields the additional infimum in condition (2.8). A similar complication is absent in the volatility case; compare Blasques et al. (2014b) and Harvey (2013).

Fourth, condition (2.8) simplifies for particular choices of parameterizations and scale functions. For example, if we use the familiar Fisher transformation $\rho(f_t;\lambda) = \tanh(f_t)$, we have $\dot{\rho}(f_t) = 1 - \rho(f_t)^2$ and the entire middle term in (2.8) vanishes. For this particular parameterization and fixing the scaling function to $S(f_t;\lambda) \equiv 1$, we can even provide further analytical results relating to the optimal choice of the function ψ. Using a Jensen, triangle, and Cauchy-Schwarz inequality, we obtain a stricter sufficient condition for SE from (2.8) as

$$\begin{aligned}
&\inf_{\psi\in\Psi} \mathbb{E}\sup_{f\in\mathcal{F}}\left|\beta + \alpha(1-\rho(f_t;\lambda)^2)\dot{g}(\rho(f_t;\lambda))k(u_t)\right| \leq \\
&\qquad |\beta| + |\alpha| \inf_{\psi\in\Psi} \mathbb{E}\sup_{f\in\mathcal{F}}\left|(1-\rho(f_t;\lambda)^2)\dot{g}(\rho(f_t;\lambda))k(u_t)\right| \leq \\
&\qquad |\beta| + |\alpha| \,\mathbb{E}\|k(u_t)\| \cdot \inf_{\psi\in\Psi}\sup_{f\in\mathcal{F}}\left\|(1-\rho(f_t;\lambda)^2)\ \dot{g}(\rho(f_t;\lambda))\right\| < 1, \qquad (2.12)
\end{aligned}$$

where $\|\cdot\|$ denotes the standard Euclidean norm. Instead of the Cauchy-Schwarz inequality, we can also use a second triangle inequality to obtain the alternative stricter sufficient condition

$$|\beta| + |\alpha| \inf_{\psi\in\Psi} \mathbb{E} \sup_{f\in\mathcal{F}} \left|(1-\rho(f_t;\lambda)^2)\dot{g}(\rho(f_t;\lambda))k(u_t)\right| \leq$$
$$|\beta| + |\alpha| \inf_{\psi\in\Psi} \sum_{i=1}^{3} \sup_{f\in\mathcal{F}} \left|(1-\rho(f_t;\lambda)^2)\ \dot{g}_i((\rho(f_t;\lambda)))\right| \cdot \mathbb{E}\,|k_i(u_t)| < 1, \quad (2.13)$$

where $\dot{g}_i$ and k_i are the ith elements of $\dot{g}$ and k, respectively. Using either of the more stringent SE conditions (2.12) or (2.13), we obtain the following result.

Lemma 2. *Under the assumptions stated in Lemma 1, setting* $\psi(\rho) = k_\psi \arcsin(\rho)$ *with* $k_\psi = 1/2$ *reaches the functional lower bound for the sufficient condition stated in either equation* (2.12) *or* (2.13)*. The condition then reduces to* $|\beta| + |\alpha|\ \mathbb{E}\|k(u_t)\| < 1$ *for condition* (2.12) *and* $|\beta| + |\alpha|\ \mathbb{E}|k_1(u_t)| < 1$ *for condition* (2.13)*, respectively, where* $k_1(u_t)$ *is the first element of* $k(u_t)$*. The link function becomes the symmetric matrix root*

$$h(f_t;\lambda) = \begin{pmatrix} \cos\big(\arcsin(\rho(f_t;\lambda))/2\big) & \sin\big(\arcsin(\rho(f_t;\lambda))/2\big) \\ \sin\big(\arcsin(\rho(f_t;\lambda))/2\big) & \cos\big(\arcsin(\rho(f_t;\lambda))/2\big) \end{pmatrix}.$$

The result in Lemma 2 shows that we uniformly obtain the largest SE region for the stricter conditions (2.12) or (2.13) for the symmetric matrix root h in (2.6). The choice of h in setting up the dynamic equation (2.6) is thus far from innocuous and

directly influences the size and shape of the SE region.

2.4 Asymptotic properties of the maximum likelihood estimator

In this section we establish the invertibility of the GAS model, as well as the consistency and asymptotic normality of the ML estimator (MLE) for the static parameters θ.

Model invertibility is a critical element in the proof of consistency and asymptotic normality of the MLE since the filter (2.2) enters the likelihood function and must thus be ensured to have appropriate stochastic properties. Similarly to Straumann and Mikosch (2006), this section uses the contraction condition of Bougerol (1993) in order to ensure model invertibility and bounded moments for the filtering sequence. This is crucial for the asymptotic properties of the MLE since the initialized time-varying parameter and its derivatives enter the likelihood function and its derivatives. The following result builds on the SE nature of the data $\{y_t\}_{t\in\mathbb{Z}}$ which follows easily from the SE nature of the true time-varying parameter $\{f_t\}_{t\in\mathbb{Z}}$ established in the previous section.

Lemma 3. (Model Invertibility) *Let Θ be compact, let $\{y_t\}_{t\in\mathbb{Z}}$ generated by (2.1)–(2.3) be SE, let the scaled score s be smooth in all arguments and Lipschitz in $f \in \mathcal{F}$,*

and assume that there exists a non-random f_1 *such that*

(i) $\mathbb{E}\log^+|S(f_1)q(f_1,y_t)|<\infty$

(ii) $\mathbb{E}\log\sup_{\theta\in\Theta}\sup_{f^*}\left|\alpha\Big(\dot{S}(f^*)q(f^*,y_t)+S(f^*)\dot{q}(f^*,y_t)\Big)+\beta\right|<1,\ f^*\in\mathcal{F}$

where $\log^+ x\equiv\max(\log x,0)$ *for* $x\in\mathbb{R}_+$ *and* $q(f_t,y_t)$ *is the score expression in (2.4). Then the GAS recursion* $\{f_t(\theta,f_1)\}_{t\in\mathbb{N}}$ *defined in (2.2) converges e.a.s. to a unique limit SE process* $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ *that admits the representation* $f_t(\theta)=\Phi(y_{t-1},y_{t-2},...)$ *for every* t *and some measurable function* Φ*.*

Note that the contraction condition (ii) in Lemma 3 is different from the one studied in Lemma 1 since it refers to the filtering equation that takes the data y_t as given. By contrast, the contraction property in Lemma 1 looks at the GAS model as a data generating process, and hence defined the data y_t in terms of the true unknown parameter f_t and the innovations u_t. In other words, while Lemma 1 deals with *simulated* data, Lemma 3 deals with *observed* data. The invertibility condition above is in particular required in order to make dependence on the fixed initial value f_1 vanish in the GAS recursion and therefore in the objective function. To make this transparent, let $\ell_t(\theta,f_1)$ denote the time t log-likelihood contribution for the vector of static parameters θ, and

$$\mathcal{L}_T(\theta):=\frac{1}{T}\sum_{t=1}^{T}\ell_t(\theta)=\frac{1}{T}\sum_{t=1}^{T}\left(\log p(y_t^\top R(f_t(\theta,f_1))y_t)-\log|R(f_t(\theta,f_1))|^{1/2}\right). \tag{2.14}$$

Define $\mathcal{L}_T(\theta)$ and $\ell_t(\theta)$ similar to (2.14), but with the limiting process $f_t(\theta)$ replacing $f_t(\theta, f_1)$.

Theorem 1 establishes the strong consistency of the MLE assuming the identification of the true parameter vector $\theta_0 \in \Theta$. The strong consistency result holds for every initialization of the filter satisfying the conditions of Lemma 3.

Theorem 1. (Consistency) *Let* $(\Theta, \mathfrak{B}(\Theta))$ *be a compact measurable space and let the conditions of Lemmas 1 and 3 hold. Furthermore, assume that* $\mathbb{E}\sup_{\theta\in\Theta}|\ell_t(\theta)| < \infty$ *and that* $\theta_0 \in \Theta$ *is the unique maximizer of* $\mathcal{L}_\infty(\theta) \equiv \mathbb{E}\ell_t(\theta)$*. Then, for every* $f_1 \in \mathcal{F}$*, the MLE, defined as* $\hat{\theta}_T(f_1) := \arg\max_{\theta\in\Theta}\mathcal{L}_T(\theta)$ *satisfies* $\hat{\theta}_T(f_1) \xrightarrow{a.s.} \theta_0$ *as* $T \to \infty$*.*

Theorem 2 establishes the asymptotic normality of the MLE. In this theorem we let $\mathcal{I}(\theta_0) := -\mathbb{E}\ddot{\ell}_t(\theta_0)$ denote the Fisher information matrix and $\mathcal{J}(\theta_0) := \mathbb{E}\dot{\ell}_t(\theta_0)\dot{\ell}_t(\theta_0)^\top$ is the expected outer product of gradients, with $\dot{\ell}_t(\theta)$ and $\ddot{\ell}_t(\theta)$ denoting first and second order derivatives of $\ell_t(\theta)$ with respect to θ, respectively.

Theorem 2. (Asymptotic Normality) *Let the conditions of Theorem 1 hold and let* θ_0 *be a point in the interior of* Θ*. Furthermore, let the first and second derivatives of the log likelihood contributions* $\ell_t(\theta)$ *be asymptotically SE and satisfy* $\mathbb{E}|\dot{\ell}_t(\theta_0)|^2 < \infty$ *and* $\mathbb{E}\sup_{\theta\in\Theta}|\ddot{\ell}_t(\theta)| < \infty$*. Then, for every* $f_1 \in \mathcal{F}$*, the ML estimator* $\hat{\theta}_T(f_1)$ *satisfies*

$$\sqrt{T}(\hat{\theta}_T(f_1) - \theta_0) \xrightarrow{d} \mathrm{N}\big(0, \mathcal{I}^{-1}(\theta_0)\mathcal{J}(\theta_0)\mathcal{I}^{-1}(\theta_0)\big) \quad as \quad T \to \infty.$$

The theoretical results in the previous two theorems are supported by unreported

simulation experiments. We find that increasing the sample size brings the ML estimates over repeated Monte Carlo replications closer to their true values in a controlled setting. Moreover, we also find that the distribution of the estimator approaches the normal distribution for increasing sample sizes.

2.5 Numerical and empirical results

2.5.1 Numerical results

Alternative choices for the error density generator p, the scaling function S, the parameterization ρ, and the matrix square root h give rise to different models with different SE regions. For a number of these choices, we check for every pair (α, β) whether the sufficient condition (2.8) is satisfied. We plot the SE region in the (α, β)-plane by numerically identifying, for every fixed β, the corresponding maximum and minimum values of α that satisfy (2.8).

We propose the following computationally efficient algorithm to speed up the computations. The maximization inside the worst case contraction condition (2.7) can be conducted in two steps – separating the effect of score from the static parameters (α, β).

Lemma 4. *The calculation of the score-driven contraction property obeys the follow-*

ing identity

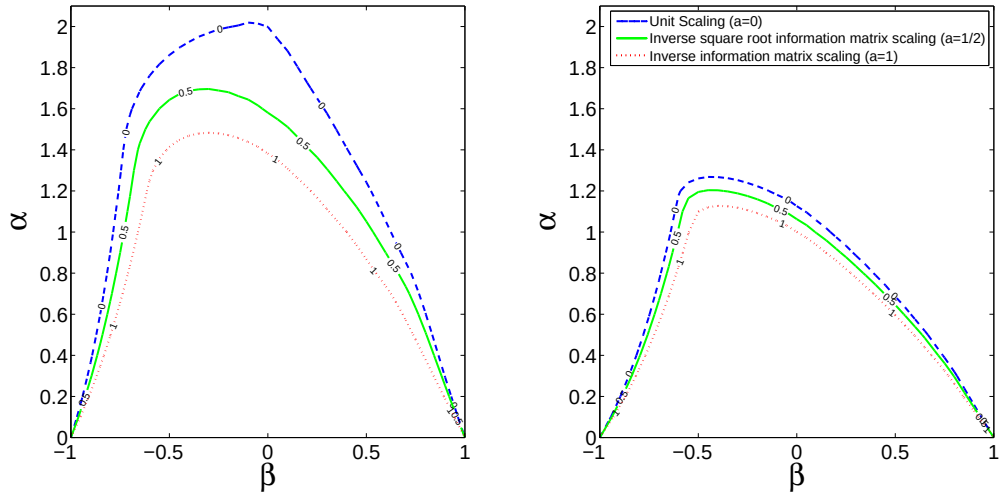
$$\mathbb{E}\log\sup_{f\in\mathcal{F}}|\beta+\alpha\dot{s}(f,\alpha,\beta,u,\psi)|=\mathbb{E}\log\max\{|\beta+\alpha\sup_{f}\dot{s}(f,u,\psi)|,|\beta+\alpha\inf_{f}\dot{s}(f,u,\psi)|\}.$$

In other words, computation of the extremal scores $\sup_f \dot{s}(f,u,\psi)$ is sufficient for obtaining the extremal scores for any other (α',β'). Even though for fixed (α,β) this representation requires twice the floating point operations as for any other (α',β') we can reuse the up and down extremal scores. The computational effort then grows linearly with the number of points considered, whereas performing the original computation for every (α,β) grows with a higher polynomial degree.

To fix ideas, consider the class of Student's t densities for u_t as in Creal et al. (2011). The Fisher transformation $\rho(f_t)=\tanh(f_t)$ ensures proper value for the correlation parameter. As indicated in Section 2.3, this also simplifies the evaluation of the SE condition in Lemma 1. For the scaling function S, we adopt the three choices based on the information matrix as presented in equation (2.5).

Next, we investigate the sensitivity of the SE region to the choice of matrix root $h(\cdot)$. For this, we consider two prominent alternatives, both described by $\psi(\rho)=k_\psi\arcsin(\rho)$ for $k_\psi\in\mathbb{R}$. The first alternative is the symmetric matrix root of Lemma 2 with $k_\psi=1/2$. The second is the familiar (lower triangular) Cholesky decomposition, which is obtained by setting $k_\psi=1$.

To numerically evaluate the sufficient SE condition (2.8), we need to solve an opti-



(a) Symmetric root ($k_\psi = 1/2$) (b) Cholesky root ($k_\psi = 1$)

Figure 2.1: Contraction sufficiency regions for the normal distribution and different scaling choices $S(f_t;\lambda) = (\mathcal{I}_t(f_t;\lambda))^{-a}$ for $a \in \{0, 1/2, 1\}$

Notes: The two panels contain different regions obtained by parameterizing the matrix roots $h(f_t;\lambda)$ with $\psi(\rho) = k_\psi \arcsin(\rho)$. Panel (a) contains the results for the symmetric matrix root ($k_\psi = 1/2$) and panel (b) corresponds to the Cholesky decomposition ($k_\psi = 1$).

mization problem within an integration procedure. As the state equation is univariate, the integral can be evaluated via a quadrature rule. We can gain further numerical efficiency for the inner optimization problem by storing maximum and minimum values of $S(f_t;\lambda)\, q(h(f_t;\lambda)u_t, f)$ for each point u_t and recycling these for evaluation at different points in the (α, β)-plane as described by Lemma 4. We avoid local optima by evaluating the function over a wide grid and by noting that for the Student's t distribution $(\partial/\partial f)^i s_t(f_t;\lambda) \to 0$ as $|f| \to \infty$ for all $i > 1$. We can further halve the computation time by noting that in our setting $|\partial\phi_t(f;\theta)/\partial f| = |\partial\phi_t(f;-\theta)/\partial f|$.

In panel (a) of Figure 2.1, we present the results for the normal distribution and the symmetric root $h(f_t; \lambda)$, i.e., $k_\psi = 1/2$. The figure contains three different regions, each one corresponding to a different form of scaling in equation (2.5). Points inside each region are combinations of (α, β) for which the sufficient condition (2.8) is met. The shape of the sufficient SE region is anti-symmetric around the origin due to the absolute signs in (2.8), such that we only plot its upper half. The region also shows a non-monotonic curvature, particularly in the second quadrant. This feature is due to the use of absolute values, the change in the location of the supremum in (2.8) in the second quadrant, and a shift in the relevant region of integration if the derivative of $S(f_t; \lambda)q(h(f_t; \lambda)u_t, f)$ changes sign.

An interesting feature in Figure 2.1 is the behavior of the region for square root inverse information matrix scaling, $a = 1/2$ in (2.5). First note that $a = 1/2$ has the property that the update via $s_t(f_t, y_t)$ is invariant with respect to reparametrizations of f_t. Furthermore, under correct specification the steps $S(f_t; \lambda)q(y_t, f)$ in (2.4) are by construction martingale differences with unit variance; see also Creal et al. (2013). This implies that $\{f_t\}_{t\in\mathbb{N}}$ converges to a covariance stationary process as long as $|\beta| < 1$. The region in Figure 2.1 shows that $|\beta| < 1$ is necessary, but not sufficient for (2.8) to be satisfied. This relates directly to discussions in the GARCH literature, where in the univariate setting covariance stationarity is a more restrictive condition than strict stationarity, but the relation between the two remains an open question in a multivariate context; see for example Boussama et al. (2011).

Panel (b) in Figure 2.1 shows the different SE regions for a different choice of matrix root $h(f_t; \lambda)$, namely the Cholesky decomposition. It is clear that the sufficiency regions in the (α, β)-plane are smaller than the corresponding regions for the symmetric root. As models constructed with a symmetric root and a Cholesky root are observationally equivalent, we can take the larger regions as sufficient regions for SE to hold; see also Lemma 2. The differences make clear that the choice of matrix root is important for determining the size of the region either analytically or numerically.

We provide more SE regions in Appendix 2.C, including regions based on the further inequalities used to establish Lemma 2. In particular, we show that the SE regions for the Student's t distribution under square root information matrix scaling ($a = 1/2$) are smaller for fatter tails if the Cholesky decomposition is used ($k_\psi = 1$), while the converse holds if we use the symmetric root decomposition ($k_\psi = 1/2$). In the Section 2.5.2 we document how this wedge may also become empirically relevant.

2.5.2 Empirical illustration

In this section we study the time-varying correlation between the London and Athens stock exchange. We take daily returns of the FTSE 100 and the Athex Composite over the period January 1, 2002 to March 2, 2013. We are particularly interested in whether there are indications that the correlation between these two markets changed over the course of the European sovereign debt crisis.

To focus on the correlation part of the model, we first filter both series using AR-

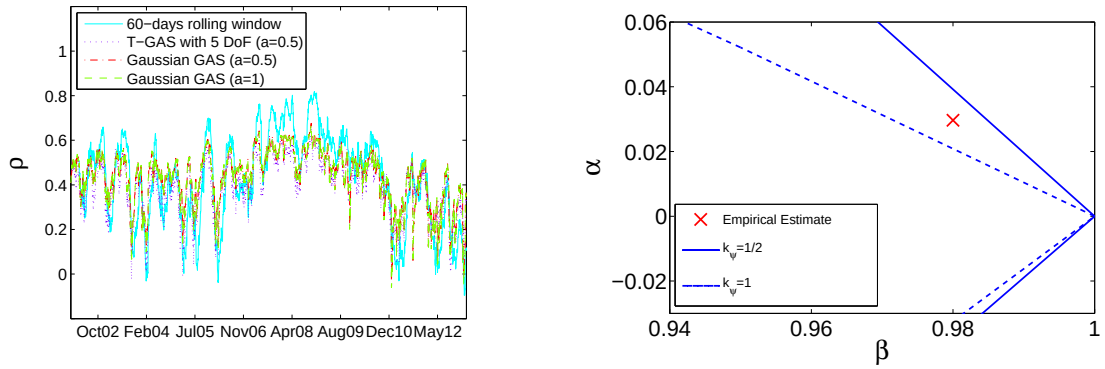
GARCH type models; see also Patton (2009). The mean of both series is modeled by a first order autoregressive progress. We find a strong leverage effects in both series and capture these by a GJR(1,2) model of Glosten et al. (1993) for the FTSE index, and an EGARCH(1,2) specification of Nelson (1991) for the Athens index, respectively.[4] We test the null hypothesis of constant and zero residual correlation against time-varying alternatives using a Nyblom test of the form

$$NB_c = \frac{1}{\hat{\sigma}^2}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{t}\left(x_{1s}x_{2s}-\bar{\rho}\right)\right)^2 \xrightarrow{d} \int_0^1 B_b(z)^2\,dz$$
$$NB_0 = \frac{1}{\hat{\sigma}_0^2}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{t}x_{1s}x_{2s}\right)^2 \xrightarrow{d} \int_0^1 B(z)^2\,dz,$$

where $B(z)$ and $B_b(z) = B(z) - zB(1)$ denote a standard Brownian motion and a standard Brownian bridge, respectively. The average correlation is estimated by $\bar{\rho} = T^{-1}\sum_{t=1}^{T} x_{1t}x_{2t}$ and $\hat{\sigma}^2$ is a heteroskedasticity and autocorrelation consistent estimator of the long-run variance of $(x_{1t}x_{2t} - \bar{\rho})$, with $\hat{\sigma}_0^2$ defined similarly when $\bar{\rho}$ is set to 0.[5] By letting x_t denote the univariate volatility-filtered series, i.e. $x_t := y_t$ in the notation of our paper, we find strong evidence for time-varying correlations. We therefore use the filtered univariate series to estimate the GAS model from Section 3.2 with a time varying correlation coefficient. The Nyblom test can further be used as a diagnostic for remaining time-variation in dynamic correlations when applied to

[4]Further robustness results for alternative specifications for the marginals can be found in Appendix 2.D.

[5]Critical values of the test are simulated by discretizing the processes $B(\cdot)$ and $B_b(\cdot)$ and can also be found in Hansen (1992).



(a) Correlation paths

(b) Student's $t(5)$ estimates ($a = 1$)

Figure 2.2: Empirical results

Notes: Panel (a) shows 60-day rolling correlations and the filtered correlations between the FTSE 100 (UK) and Athex Composite (Greece) equity index returns. Panel (b) on the right puts the empirical estimates obtained by unconstrained estimation into the zoomed stationarity and ergodicity region perspective. The red cross depicts the empirical point estimate of the model, while the solid (dashed) line marks the contraction region when using symmetric (Cholesky) decomposition of the correlation matrix.

estimated residuals $x_t := \hat{u}_t = h^{-1}(\hat{f}_t)y_t$. The results are shown in Figure 2.2 and Table 2.1.

Panel (a) in Figure 2.2 shows the dynamic correlations between the filtered series. As a non-parametric benchmark, we plot simple 60-day rolling window correlations. The rolling window estimates suggest that correlations exhibit clear signs of time variation. Correlations lie around 0.4 up to about 2006, then increase to about 0.6, and come down substantially to around 0.2 during the sovereign debt crisis. On top of this slow variation, there are also substantial dynamic patters at higher frequencies.

The possibly lower correlations between the UK and Greek stock indices are inter-

esting economically, particularly given the stable correlation pattern between the two series during almost the whole of the prelude, height, and aftermath of the preceding financial crisis (2007–2009). The financial crisis, apparently, did not substantially alter the real economic linkages between the two economies as evidenced by the stable dynamic of the correlations between the two stock markets. It is only after the announcement of the record Greek deficit late 2009 and the subsequent actions that gave rise to the European sovereign debt crisis, that the link between the euro denominated Greek stock market and the sterling denominated FTSE is broken. The correlations remain at low levels even after the non-standard monetary policy actions by the European Central Bank.

Table 2.1 provides the parameter estimates of the GAS models. We provide a benchmark by estimating a simple exponentially weighted moving average (EWMA) scheme for the correlation based on the recursion $\rho_t = \tanh(f_t)$ and

$$f_{t+1} = \omega + \beta f_t + (1 - \beta) y_{1t} y_{2t},$$

see also the Gaussian dynamic copula specification of Patton (2006b).

We see that the GAS model is empirically useful both in terms of in-sample likelihood fit and improving the diagnostics for time-varying correlation. All models indicate that the correlation parameter is highly persistent: the estimated values of β all lie very close to 1. The scaling function for the score only has a mild effect on the

Table 2.1: Estimation Results

Notes: The table reports estimation output for the models discussed in Section 2.5.2. For the GAS model specifications, we also compute whether the contraction property holds for symmetric and Cholesky decompositions of the correlation matrix. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are in parentheses. Best log-likelihood, AIC, and BIC values across models are printed in boldface. $H_0^{5\%}$: $\rho_t \equiv 0$ and $H_0^{5\%}$: $\rho_t \equiv \bar{\rho}$ indicate whether the Nyblom test among residuals rejects constant zero or estimated correlation, respectively. HAC standard errors are computed using Newey-West weights with $\min(\lfloor 1.2 \times T^{1/3} \rfloor, T)$ lags.

	EWMA	EWMA	EWMA	t(∞)-GAS ($a=0$)	t(∞)-GAS ($a=1/2$)	t(∞)-GAS ($a=1$)	t(5)-GAS ($a=0$)	t(5)-GAS ($a=1/2$)	t(5)-GAS ($a=1$)	t(λ)-GAS ($a=1$)
λ	5	5	8.9321	∞	∞	∞	5	5	5	9.1411
			(0.0801)							(0.9287)
ω		0.0003	0.0012	0.0117	0.0115	0.0112	0.0089	0.0089	0.0088	0.0114
		(0.0003)	(0.0003)	(0.0016)	(0.0014)	(0.0013)	(0.0011)	(0.0011)	(0.0010)	(0.0013)
α				0.0254	0.0283	0.0314	0.0335	0.0316	0.0297	0.0360
				(0.0040)	(0.0045)	(0.0050)	(0.0048)	(0.0046)	(0.0043)	(0.0050)
β	0.9757	0.9756	0.9739	0.9752	0.9757	0.9763	0.9797	0.9798	0.9800	0.9762
	(0.0085)	(0.0063)	(0.0030)	(0.0021)	(0.0016)	(0.0012)	(0.0013)	(0.0011)	(0.0009)	(0.0012)
Log-likelihood	−7847	−7847	−7812	−7891	−7890	−7890	−7845	−7845	−7845	**−7809**
AIC	15697	15699	15631	15778	15787	15787	15696	15696	15696	**15626**
BIC	15703	15711	**15649**	15814	15805	15805	15714	15714	15714	**15649**
# parameters	1	2	3	3	3	3	3	3	3	4
$H_0^{5\%}$: $\rho_t \equiv 0$	reject	reject					reject	reject	reject	
$H_0^{5\%}$: $\rho_t \equiv \bar{\rho}$		reject	reject	reject	reject	reject	reject	reject	reject	
Inside SE region?										
Cholesky ($k_\psi = 1$)				Yes	Yes	No	No	No	No	No
Symmetric ($k_\psi = \frac{1}{2}$)				Yes	Yes	Yes	Yes	Yes	Yes	Yes

model's fit: the likelihood values are similar for every $a \in \{0, 1/2, 1\}$. The degrees of freedom parameter λ is estimated around 9. This is substantially fatter-tailed than the normal, but also substantially lighter-tailed than the Student's t distribution with the degrees of freedom fixed at 5. The differences in likelihood values, as well as Akaike (AIC) and Bayes (BIC) information criteria indicate that estimating the degrees of freedom improves the fit of the model significantly. Furthermore, the estimation of the tail parameter λ also improves the fit in terms of model diagnostics: only the GAS model with estimated degrees of freedom passes the Nyblom tests for remaining time-variation in the residual cross-correlations.

We plot the SE region boundaries for the Cholesky and the symmetric root decomposition in panel (b) of Figure 2.2. The estimated values of α and β for the $t(\lambda)$-GAS specification are indicated by the cross mark. The results clearly confirm the importance of the choice of ψ in verifying the SE properties. For the symmetric root based region, we obtain stationarity and ergodicity at the estimated parameter values. For the Cholesky decomposition, by contrast, we fail to obtain this result. As condition (2.8) takes the infimum over ψ and thus the widest region in panel (b) over all admissible decompositions $h(f_t)$, the Cholesky decomposition is in this setting less powerful to discriminate SE from non-SE regions of the parameter space. We stress again that all of these regions are only based on sufficient conditions, and that the actual regions may be wider.

For all models considered, the bottom lines in Table 2.1 indicate whether the

estimated parameters lie inside the SE region. For the Gaussian models with unit ($a = 0$) and inverse square root information matrix scaling ($a = 1/2$) the choice of matrix decomposition does not have an impact. Both the symmetric root ($k_\psi = 1/2$ in (2.11)) and Cholesky root ($k_\psi = 1$ in (2.11)) indicate that the estimated parameters are inside the SE region and satisfy the average contraction condition. For inverse information matrix scaling, however, and for the Student's t based models, we find a similar difference as in panel (b) of Figure 2.2: we cannot ensure SE properties based on the Cholesky decomposition, whereas we can do so for the symmetric root. This again highlights that the use of different constructive devices such as different matrix decompositions is empirically relevant for the verification of sufficient SE conditions in a multivariate setting.

2.6 Concluding remarks

We have derived sufficient regions for stationarity and ergodicity for a new class of score driven dynamic correlation models. The regions exhibit a highly non-standard shape. Moreover, we have shown that the shape and size of the SE regions depends on the type of matrix root that is chosen in checking the sufficient conditions of Bougerol (1993). Furthermore, we have seen how the stationarity conditions can be used in establishing results for consistency and asymptotic normality of the maximum likelihood estimator. The numerical stability conditions were supported by an empirical

investigation of the time varying correlation between UK and Greek stock markets. We found a substantial drop in the linkages between the sterling denominated UK market and the euro denominated Greek market over the course of the European sovereign debt crisis. Such a break in dependence between markets, however, was absent during the preceding global financial crisis.

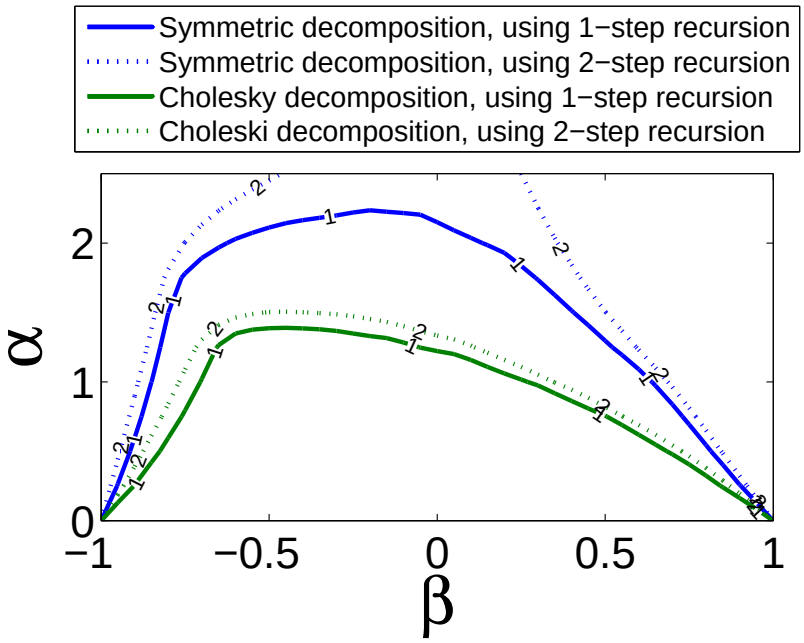


Figure 2.3: Comparison of worst case contractions for different recursion step lengths

Notes: Gaussian model with unit scaling.

Looking forward, we are further interested in establishing sharp properties that go beyond the worst case single-step limiting recursions. By considering a multi-step recursion in Figure 2.3, one can relax the sufficiency regions at the expense of a painfully expensive computational burden. An interesting possible extension of

our current results concerns a generalization to the fully multivariate (rather than bivariate) setting of score driven correlation models proposed in Creal et al. (2011). Due to the difficulties of the contraction condition under multivariate time-varying parameters as pointed out by (Boussama, 1998, p. 131), we will follow a distributional approach, as opposed to the pathwise properties explored in the current chapter. This enables us to give sharper bounds for stationarity and ergodicity in the context of even very high-dimensional dynamic correlation models.

2.A Proofs

We first state Theorem 3.1 of Bougerol (1993). Denote by $\log \Lambda(\phi_0)$ the term inside the expectation on the left hand side of (2.7).

Theorem 3 (Bougerol (1993, Theorem 3.1))**.** *Let $\{\phi_t\}$ be a stationary and ergodic sequence of Lipschitz maps $\phi_t : \mathcal{F} \to \mathcal{F}$. Assume*

1. *There exists a $f \in \mathcal{F}$ and distance measure d such that $\mathbb{E}[\log^+ d(\phi_0(f), f)] < \infty$;*
2. $\mathbb{E}[\log^+ \Lambda(\phi_0)] < \infty$;
3. $\mathbb{E}[\log \Lambda(\phi_0^{(r)})] < 0$, *where $\phi_0^{(r)}$ denotes the r-fold backward iterates.*

Then the stochastic recurrence equation (2.6) admits a unique stationary ergodic solution $\{f_t\}$.

Proof of Lemma 1. The SE property of $\{f_t\}$ follows from the measurability with respect to $\{u_t\}$. The Lipschitz property is obtained from the boundedness of the terms in equation (2.20) below. Condition 1 is then ensured by the definition of the GAS transition equation and the assumed moments in Lemma 1, as we can write $\mathbb{E}[\log^+ d(\phi_0(f), f)] \leq \mathbb{E}|\phi_0(f) - f| = \mathbb{E}|\omega + (\beta - 1)f + \alpha S(f)q(h(f)u_0, f)| \leq |\omega| + |(\beta - 1)f| + \alpha \mathbb{E}|S(f)q(h(f)u_0, f)|$. As requirement 2 is implied by 3, we can now turn our main interest towards the study of the latter, non-trivial, condition 3.

We write ξ and ψ for $\xi(\rho)$ and $\psi(\rho)$, respectively. Define the shorthand notation $c_w = c_w(\rho) = \cos(w(\rho))$ with $w : [-\delta, \delta] \to \mathbb{R}$, and similarly $s_w = s_w(\rho) = \sin(w(\rho))$. Each matrix root h of the correlation matrix can be written as

$$h(f_t;\lambda) = \begin{pmatrix} c_\xi(\rho(f_t;\lambda)) & s_\xi(\rho(f_t;\lambda)) \\ s_\psi(\rho(f_t;\lambda)) & c_\psi(\rho(f_t;\lambda)) \end{pmatrix}. \tag{2.15}$$

Using (2.15), we obtain

$$\begin{pmatrix} c_\xi & s_\xi \\ s_\psi & c_\psi \end{pmatrix} \begin{pmatrix} c_\xi & s_\psi \\ s_\xi & c_\psi \end{pmatrix} = \begin{pmatrix} 1 & s_{\xi+\psi} \\ s_{\xi+\psi} & 1 \end{pmatrix},$$

such that we require $\sin(\psi + \xi) = \rho$ or $\xi(\rho) = \arcsin(\rho) - \psi(\rho)$ for some arbitrary function $\psi(\rho)$. It follows that $s_\xi = \rho c_\psi - \sqrt{1-\rho^2} s_\psi$, and $c_\xi = \sqrt{1-\rho^2} c_\psi + \rho s_\psi$. From this we obtain

$$\begin{pmatrix} c_\xi & s_\psi \\ s_\xi & c_\psi \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_\xi & s_\xi \\ s_\psi & c_\psi \end{pmatrix} =$$

$$\begin{pmatrix} 2c_\xi s_\psi & s_\xi s_\psi + c_\xi c_\psi \\ s_\xi s_\psi + c_\xi c_\psi & 2 s_\xi c_\psi \end{pmatrix} = \begin{pmatrix} 2c_\xi s_\psi & c_{\xi-\psi} \\ c_{\xi-\psi} & 2 s_\xi c_\psi \end{pmatrix} =$$

$$\begin{pmatrix} -\rho c_{2\psi} + \sqrt{1-\rho^2} s_{2\psi} + \rho & \sqrt{1-\rho^2} c_{2\psi} + \rho s_{2\psi} \\ \sqrt{1-\rho^2} c_{2\psi} + \rho s_{2\psi} & \rho c_{2\psi} - \sqrt{1-\rho^2} s_{2\psi} + \rho \end{pmatrix} =: H(\rho) + \rho \mathrm{I}.$$

Using $y_t = h(f_t;\lambda)u_t$, we can rewrite (2.4) as

$$
\begin{aligned}
(1-\rho(f_t;\lambda)^2)q(h(f_t;\lambda)u_t, f)/\dot{\rho}(f_t;\lambda) = \\
& 2\dot{p}(u_t^\top u_t)\rho(f_t;\lambda)u_t^\top u_t - \dot{p}(u_t^\top u_t)u_t^\top \left(H(\rho(f_t;\lambda)) + \rho(f_t;\lambda)\mathrm{I}\right)u_t + \rho(f_t;\lambda) = \\
& \dot{p}(u_t^\top u_t)\rho(f_t;\lambda)u_t^\top u_t - \dot{p}(u_t^\top u_t)u_t^\top H(\rho(f_t;\lambda))u_t + \rho(f_t;\lambda) = \\
& \rho(f_t;\lambda)\left(\dot{p}(u_t^\top u_t)u_t^\top u_t + 1\right) - \dot{p}(u_t^\top u_t)u_t^\top H(\rho(f_t;\lambda))u_t = g(\rho)k(u_t),
\end{aligned}
\tag{2.16}
$$

with

$$
g(\rho) = \left(\rho\ ,\ \rho c_{2\psi} - \sqrt{1-\rho^2}s_{2\psi}\ ,\ \sqrt{1-\rho^2}c_{2\psi} + \rho s_{2\psi}\right),
$$

$$
k(u_t) = \left(\dot{p}(u_t^\top u_t)u_t^\top u_t + 1\ ,\ \dot{p}(u_t^\top u_t)(u_{1,t}^2 - u_{2,t}^2)\ ,\ -2\dot{p}(u_t^\top u_t)u_{1,t}u_{2,t}\right)^\top,
$$

and $u_t = (u_{1,t}, u_{2,t})^\top$. Defining $\dot{g}(\rho) = \partial g(\rho)/\rho$ as the derivative of $g(\rho)$, it holds that

$$
\begin{aligned}
\dot{g}(\rho) &= \left(1\ ,\ c_{2\psi}(\rho) + \rho\cdot(1-\rho^2)^{-1/2}s_{2\psi}(\rho)\ ,\ -\rho\cdot(1-\rho^2)^{-1/2}c_{2\psi}(\rho) + s_{2\psi}(\rho)\right) + \\
&\quad 2\dot{\psi}(\rho)\left(0\ ,\ -\rho s_{2\psi}(\rho) - \sqrt{1-\rho^2}c_{2\psi}(\rho)\ ,\ -\sqrt{1-\rho^2}s_{2\psi}(\rho) + \rho c_{2\psi}(\rho)\right) \qquad (2.17) \\
&= (1,0,0) + \left((1-\rho^2)^{-1/2} - 2\dot{\psi}(\rho)\right)\cdot \\
&\quad \left(0\ ,\ \sqrt{1-\rho^2}c_{2\psi}(\rho) + \rho\cdot s_{2\psi}(\rho)\ ,\ \sqrt{1-\rho^2}s_{2\psi}(\rho) - \rho\cdot c_{2\psi}(\rho)\right). \qquad (2.18)
\end{aligned}
$$

The definitions in (2.6) and (2.16) then imply that (2.7) can be written as

$$\frac{\partial \phi_t(f;\theta)}{\partial f} = \beta + \alpha \left(\frac{\partial}{\partial f} \left(\frac{S(f_t;\lambda)\ \dot{\rho}(f_t;\lambda)}{1-\rho(f_t;\lambda)^2} \right) \right) g(\rho(f_t;\lambda))k(u_t) + \alpha \frac{S(f_t;\lambda)\ \dot{\rho}(f_t;\lambda)}{1-\rho(f_t;\lambda)^2} \frac{\partial g(\rho(f_t;\lambda))}{\partial f} k(u_t). \tag{2.19}$$

□

Proof of Lemma 2. Using (2.18), we can rewrite $\|\dot{g}(\rho)\|^2$ as

$$1 + \left| \frac{1}{\sqrt{1-\rho^2}} - 2\dot{\psi}(\rho) \right|^2 \cdot \left(c_{2\psi}(\rho)^2 + s_{2\psi}(\rho)^2 \right) = 1 + \left| \frac{1}{\sqrt{1-\rho^2}} - 2\dot{\psi}(\rho) \right|^2 \geq 1. \tag{2.20}$$

For $\psi(\rho) = \arcsin(\rho)/2$ the second term vanishes and we obtain the functional lower bound $(1-\rho^2)\cdot\|\dot{g}(\rho)\|^2 = 1-\rho^2$, which reaches its supremum of 1 at $\rho = 0$. The rest of the result follows directly from the definition of $\xi(\rho) = \arcsin(\rho) - \psi(\rho) = \arcsin(\rho)/2$. For computational reasons, it may be useful to note that

$$\|k(u_t)\|^2 = 2\dot{p}(u_t^\top u_t)^2 (u_t^\top u_t)^2 + 2\dot{p}(u_t^\top u_t)(u_t^\top u_t) + 1,$$

which only depends on the quadratic form $u_t^\top u_t$.

An analogous line of reasoning holds for condition (2.13) based on applying the triangle inequality twice.

□

Proof of Lemma 3. Follows immediately from Theorem 2.8 in Straumann and Mikosch (2006) by noting that our conditions (i) and (ii) imply conditions S.1 and S.2 in their theorem. □

Proof of Theorem 1. We follow Blasques et al. (2014a) and appeal to Theorem 3.13 of White (1994), and obtain $\hat{\theta}_T(f_1) \xrightarrow{a.s.} \theta_0$ from the uniform convergence of the criterion function and the identifiable uniqueness of the maximizer $\theta_0 \in \Theta$ defined e.g. in White (1994).

Existence: Note that $\mathcal{L}_T(\theta, f_1)$ is a.s. continuous in $\theta \in \Theta$ if each likelihood contribution is. This is obtained by the smoothness of the scaled score $s_t : \mathcal{F} \times \mathcal{Y} \times \Theta \to \mathbb{R}$ and the resulting continuity of f_t in θ as a composition of t continuous maps. Due to the compactness of Θ, by Weierstraß theorem the arg max set of the likelihood is non-empty a.s. and hence $\hat{\theta}_T$ exists.

Uniform convergence: By an application of the triangle inequality we have

$$\sup_{\theta \in \Theta} |\mathcal{L}_T(\theta, f_1) - \mathcal{L}_\infty(\theta)| \leq \sup_{\theta \in \Theta} |\mathcal{L}_T(\theta, f_1) - \mathcal{L}_T(\theta)| + \sup_{\theta \in \Theta} |\mathcal{L}_T(\theta) - \mathcal{L}_\infty(\theta)|.$$

The first term in (2.21) vanishes by the convergence of $f_t(y^{1:t-1}, \theta, f_1)$ to $f_t(y^{t-1}, \theta)$ which is established in Lemma 3. The maintained smoothness assumption on the scaled score ensures that $\ell_t(\cdot, f_1) = \ell(f_t(y^{1:t-1}, \cdot, f_1), y_t, \cdot)$ is continuous in $(f_t(y^{1:t-1}, \cdot, f_1), y_t)$. There thus exists a unique SE sequence $\{f_t(y^{1:t-1}, \cdot)\}_{t \in \mathbb{Z}}$ satisfying $\sup_{\theta \in \Theta} |f_t(y^{1:t-1}, \theta, f_1) - f_t(y^{t-1}, \theta)| \xrightarrow{e.a.s.} 0 \ \forall f_1 \in \mathcal{F}$. It thus follows that $\sup_{\theta \in \Theta} |\mathcal{L}_T(\theta, f_1) - \mathcal{L}_T(\theta)| \xrightarrow{a.s.} 0$ as

$t \to \infty$ by application of the continuous mapping theorem (see also Theorem 2.3[i] in Van der Vaart (2000)) for $\ell : \mathbb{C}(\Theta, \mathcal{F}) \times \mathcal{Y} \times \Theta \to \mathbb{R}$.

The second term in (2.21) vanishes by an application of the ergodic theorem (Theorem 2.7 in Straumann and Mikosch (2006)) to the sequence $\{\mathcal{L}_T(\cdot)\}$ with elements taking values in $\mathbb{C}(\Theta)$, so that $\sup_{\theta\in\Theta} |\mathcal{L}_T(\theta) - \mathcal{L}_\infty(\theta)| \overset{a.s.}{\to} 0$ as $T \to \infty$. This is obtained under the moment assumption $\mathbb{E}\sup_{\theta\in\Theta} |\ell_t(\theta)| < \infty$, by the SE nature of the sequence $\{\ell_t\}_{t\in\mathbb{Z}}$, which is implied by continuity of ℓ on the SE sequence $\{(f_t(y^{t-1}, \cdot), y_t)\}_{t\in\mathbb{Z}}$, which is SE using Lemmas 1 and 3 and Proposition 4.3 in Krengel (1985).

Identifiable uniqueness: Identifiable uniqueness of $\theta_0 \in \Theta$; i.e. $\sup_{\theta:\|\theta-\theta_0\|>\epsilon} \ell_\infty(\theta) < \ell_\infty(\theta_0)$ for all $\epsilon > 0$, follows by the assumed uniqueness of θ_0, the compactness of the parameter space Θ, and the continuity of $\mathbb{E}\ell_t(\theta)$ in $\theta \in \Theta$, which is implied by the continuity of $\mathcal{L}_T$ in $\theta \in \Theta$ and the uniform convergence of the objective function proved above; see e.g. White (1994). □

Proof of Theorem 2. We make use of the asymptotic normality conditions found e.g. in Theorem 6.4 of White (1994)). These conditions are: (i) the strong consistency of $\hat{\theta}_T \overset{a.s.}{\to} \theta_0 \in \text{int}(\Theta)$; (ii) the a.s. twice continuous differentiability of $\mathcal{L}_T(\theta, f_1)$ in $\theta \in \Theta$; (iii) the asymptotic normality of the score

$$\sqrt{T}\dot{\ell}_t\big(\theta_0, \boldsymbol{f}_1^{(0:1)}\big) \overset{d}{\to} \text{N}\big(0, \mathcal{J}(\theta_0)\big), \qquad \mathcal{J}(\theta_0) = \mathbb{E}\big(\dot{\ell}_t\big(\theta_0\big)\dot{\ell}_t\big(\theta_0\big)^\top\big), \tag{2.21}$$

where $\{\boldsymbol{f}_t^{(i)}(\theta; \boldsymbol{f}_1^{(0:i)})\}$ denotes the i-th derivative process and $\boldsymbol{f}_1^{(0:i)}$ denotes the initial conditions for the first i derivative processes, while $\{\boldsymbol{f}_t^{(i)}(\theta)\}$ denotes its asymptotic SE counterpart. (iv) the uniform convergence of the second derivative of the likelihood function

$$\sup_{\theta\in\Theta} \left\| \ddot{\mathcal{L}}_T(\theta, \boldsymbol{f}_1^{(0:2)}) - \ddot{\mathcal{L}}_\infty(\theta) \right\| \xrightarrow{a.s.} 0; \tag{2.22}$$

and (v) the non-singularity of $\ddot{\mathcal{L}}_\infty(\theta) = \mathbb{E}\ddot{\mathcal{L}}_T(\theta) = \mathcal{I}(\theta)$.

Weak Convergence of the Score: The score sequence $\{\dot{\ell}_t(\theta, f_1)\}$ depends not only on the data $\{y_t\}$ and the initialized process $\{f_t(\theta, f_1)\}$ but also on the derivative processes $\{\dot{f}_t(\theta, f_1)\} \equiv \{\partial f_t(\theta, f_1)/\partial\theta\}$. As such, the limit SE nature of the score and its smoothness properties imply that $\dot{\ell}_t(\theta, f_1) = \dot{\ell}\Big(y_t, f_t(\theta, f_1), \dot{f}_t(\theta, \dot{f}_1)\Big)$ is a continuous function of the limit SE process $\Big(y_t, f_t(\theta, f_1), \dot{f}_t(\theta, \dot{f}_1)\Big)$ and thus SE by Theorem 36.4 in Billingsley (1995). Note that the data $\{y_t\}$ is SE under the conditions of Lemma 1, and the process $\{f_t(\theta, f_1)\}$ and its derivative $\{\dot{f}_t(\theta, \dot{f}_1)\}$ both converge e.a.s. to an SE limit under the conditions of Lemma 3 since it is easy to show that the contraction condition in (ii) of Lemma 3 for $\{f_t(\theta, f_1)\}$ is also the relevant contraction condition for any derivative process $\{f_t^{(i)}(\theta, f_1^{(i)})\}$ of any order; see Blasques et al. (2014a).

The remainder of the proof now follows along similar lines as in Blasques et al. (2014a, Theorem 4). As a continuous function of the SE process $\{y_t, f_t(\theta), \dot{f}_t(\theta)\}$, the score sequence $\{\dot{\ell}_t(\theta)\}$ is also SE and we can apply the CLT for SE martingales in

Billingsley (1961) to obtain

$$\sqrt{T}\dot{\mathcal{L}}_T(\theta_0) \xrightarrow{d} \mathrm{N}(0, \mathcal{J}(\theta_0)) \quad \text{as} \quad T \to \infty. \tag{2.23}$$

As a result we can also conclude by Theorem 18.10[iv] in Van der Vaart (2000) that

$$\sqrt{T}\dot{\mathcal{L}}_T(\theta_0, \boldsymbol{f}_1^{(0:1)}) \xrightarrow{d} \mathrm{N}(0, \mathcal{J}(\theta_0)) \quad \text{as} \quad T \to \infty,$$

if we show that

$$\|\dot{\mathcal{L}}_T(\theta_0, \boldsymbol{f}_1^{(0:1)}) - \dot{\mathcal{L}}_T(\theta_0)\| \xrightarrow{e.a.s.} 0 \quad \text{as} \quad T \to \infty, \tag{2.24}$$

since the exponential rate in (2.24) implies that $\sqrt{T}\|\dot{\mathcal{L}}_T(\theta_0, \boldsymbol{f}_1^{(0:1)}) - \dot{\mathcal{L}}_T(\theta_0)\| \xrightarrow{a.s.} 0$ as $T \to \infty$.

To establish the e.a.s. convergence in (2.24), we use the e.a.s. convergence $|f_t(y^{1:t-1}, \theta_0, f_1) - f_t(y^{t-1}, \theta_0)| \xrightarrow{e.a.s.} 0$ and $\|\boldsymbol{f}_t^{(1)}(y^{1:t-1}, \theta_0, \boldsymbol{f}_1^{(0:1)}) - \boldsymbol{f}_t^{(1)}(y^{1:t-1}, \theta_0)\| \xrightarrow{e.a.s.} 0$, as implied by the conditions of Lemma 3. From the differentiability of $\dot{\ell}_t(\theta, \boldsymbol{f}_1^{(0:1)}) = \dot{\ell}\big(\theta, y^{1:t}, \boldsymbol{f}_t^{(0:1)}(y^{1:t-1}, \theta, \boldsymbol{f}_1^{(0:1)})\big)$ in $\boldsymbol{f}_t^{(0:1)}(y^{1:t-1}, \theta, \boldsymbol{f}_1^{(0:1)})$ and the convexity of $\mathcal{F}$, we use the mean-value theorem to obtain

$$\begin{aligned}\|\dot{\mathcal{L}}_T(\theta_0, \boldsymbol{f}_1^{(0:1)}) - \dot{\mathcal{L}}_T(\theta_0)\| &\leq \sum_{j=1}^{1+n_\theta} \Big|\frac{\partial \dot{\ell}(y^{1:T}, \hat{\boldsymbol{f}}_T^{(0:1)})}{\partial f_j}\Big| \cdot \big|\boldsymbol{f}_{j,T}^{(0:1)}(y^{1:T-1}, \theta_0, \boldsymbol{f}_1^{(0:1)}) - \boldsymbol{f}_{j,t}^{(0:1)}(y^{1:T-1}, \theta_0)\big| \\ &= \sum_{i=1}^{1+n_\theta} O_p(1) o_{e.a.s}(1) = o_{e.a.s.}(1),\end{aligned}$$

where $n_\theta = \dim(\Theta)$ denotes the dimension of the static parameter vector θ and $\boldsymbol{f}_{j,t}^{(0:1)}$ denotes the j-th element of $\boldsymbol{f}_t^{(0:1)}$, and $\hat{\boldsymbol{f}}^{(0:1)}$ is on the segment connecting $\boldsymbol{f}_{j,t}^{(0:1)}(y^{1:t-1}, \theta_0, \boldsymbol{f}_1^{(0:1)})$ and $\boldsymbol{f}_{j,t}^{(0:1)}$ Note that $\boldsymbol{f}_t^{(0:1)} \in \mathbb{R}^{1+n_\theta}$ because it contains $f_t \in \mathbb{R}$ as well as $\boldsymbol{f}_t^{(1)} \in \mathbb{R}^{n_\theta}$. Finally, the last equality follows from the assumed finite moments of the likelihood derivatives and the e.a.s. convergence of the initialized process (see also (Van der Vaart, 2000, p.12)).

Uniform convergence of second derivatives: We use the triangle inequality to write

$$\sup_{\theta\in\Theta} \|\ddot{\mathcal{L}}_T(\theta, f_1) - \ddot{\mathcal{L}}_\infty(\theta)\| \leq \sup_{\theta\in\Theta} \|\ddot{\mathcal{L}}_T(\theta, f_1) - \ddot{\mathcal{L}}_T(\theta)\| + \sup_{\theta\in\Theta} \|\ddot{\mathcal{L}}_T(\theta) - \ddot{\mathcal{L}}_\infty(\theta)\|. \quad (2.25)$$

The first term vanishes a.s. with $T \to \infty$ by application of a continuous mapping theorem because the maintained smoothness assumptions ensure that $\ddot{\mathcal{L}}_T(\cdot, f_1)$ is continuous in its arguments $\{(y_t, \boldsymbol{f}_t^{(0:2)}(y^{1:t-1}, \cdot, \boldsymbol{f}_{0:2}))\}$ and the invertibility conditions of Lemma 3 guarantee that there exists a unique SE sequence $\{\boldsymbol{f}_t^{(0:2)}(y^{t-1}, \cdot)\}_{t\in\mathbb{Z}}$ such that $\sup_{\theta\in\Theta} \left\| \left(y_T, \boldsymbol{f}_T^{(0:2)}(y^{1:T-1}, \theta, \boldsymbol{f}_{0:2})\right) - \left(y_T, \boldsymbol{f}_T^{(0:2)}(y^{1:T-1}, \theta)\right) \right\| \overset{a.s.}{\to} 0$. The second term in (2.25) converges under a uniform law of large numbers by the maintained assumption that $\mathbb{E}\sup_{\theta\in\Theta} \|\ddot{\ell}_t(\theta)\| < \infty$ and the SE nature of $\{\ddot{\ell}_t\}_{t\in\mathbb{Z}}$.

Finally, the non-singularity of the limit $\ddot{\mathcal{L}}_\infty(\theta) = \mathbb{E}\ddot{\ell}_t(\theta) = \mathcal{I}(\theta)$ in (v) is implied by the uniqueness of θ_0 as a maximizer of $\ddot{\mathcal{L}}_\infty(\theta)$ in Θ. □

2.B Student's t updating recursion driven by i.i.d. noise

We consider the general model $y_t \sim p(y_t), y_t = h(f_t)u_t, h(f_t) \in \mathbb{R}^{n,n}$ and $u_t \sim p(u_t)$ i.i.d., which implies the relationship $p(y_t) = |h(f_t)|^{-1}p(h(f_t)^{-1}u_t)$. We are able to model the time-variation in dependence by parameterizing L_t in terms the dynamic factors $\rho(f_t)$.

The most general distributional form we consider is the multivariate Student's t distribution, the density of which defined by

$$p(\mathbf{y}_t|\nu, \Sigma_t) = \frac{\Gamma\left[(\nu+k)/2\right]}{\Gamma(\nu/2)[(\nu-2)\pi]^{k/2}\left|\mathbf{\Sigma}_t\right|^{1/2}}\left[1+\frac{1}{\nu-2}\mathbf{y}_t'\mathbf{\Sigma}_t{}^{-1}\mathbf{y}_t\right]^{-(\nu+k)/2}, \tag{2.26}$$

which has the additional closure property of y_t and u_t being in the same class of distributions. This definition of the t-density implies $E[\mathbf{y}_t] = 0$ and $\text{Var}[\mathbf{y}_t] = \Sigma_t$, i.e. the shape parameter ν affects only the tail thickness without having a direct influence on the variance.

We consider a multivariate Student's t density in equation (2.26). Theorem 1 in Creal et al. (2011) gives the following expression for the information matrix,

$$\mathcal{I}_{\rho,t} = \frac{1}{(\nu+4)(1-\rho_t^2)^2}\left((\nu+2)(1+\rho_t^2) - 2\rho_t^2\right), \tag{2.27}$$

and the score

$$\nabla_{\rho,t} = \frac{1}{(1-\rho_t^2)^2}\frac{1}{\nu-2+\epsilon_t^2+\eta_t^2}\Big\{\nu\left((1+\rho_t^2)(y_{1t}y_{2t}-\rho_t)-\rho_t(y_{1t}^2+y_{2t}^2-2)\right)+$$
$$(1+\rho_t^2)\left(2y_{1t}y_{2t}-\rho_t(\epsilon_t^2+\eta_t^2-2)\right)-2\rho_t\left(y_{1t}^2+y_{1t}^2-(\epsilon_t^2+\eta_t^2-2)\right)\Big\}. \quad (2.28)$$

Next we write the score entirely in terms of the independent noise $u_t = (\epsilon_t, \eta_t)^\top$ such that $\mathbb{E}[u_t u_t^\top] = \mathrm{I}_n$. However this decomposition is not unique. Two prominent choices are:

1. Cholesky root, obtained by setting $\psi(\rho) = \arcsin(\rho)$ in equation (12) of the paper:

$$h(f_t;\lambda) = \begin{pmatrix} 1 & 0 \\ \rho(f_t;\lambda) & \sqrt{1-\rho(f_t;\lambda)^2} \end{pmatrix},$$

$$\nabla_{\rho,t} = \frac{1}{1-\rho_t^2}\frac{1}{\nu-2+\epsilon_t^2+\eta_t^2}\left\{\nu\left[\sqrt{1-\rho_t^2}\epsilon_t\eta_t-\rho_t(\eta_t^2-1)\right]+\right.$$
$$\left.2\left[\sqrt{1-\rho_t^2}\epsilon_t\eta_t+\rho_t\left(\frac{1}{2}(\epsilon_t^2-\eta_t^2)-1\right)\right]\right\}.$$

2. Symmetric root, obtained by setting $\psi(\rho) = 1/2\arcsin(\rho)$ in equation (12) of the paper:

$$h(f_t;\lambda) = \begin{pmatrix} \frac{1}{2}(\sqrt{1+\rho(f_t;\lambda)} + \sqrt{1-\rho(f_t;\lambda)}) & \frac{1}{2}(\sqrt{1+\rho(f_t;\lambda)} - \sqrt{1-\rho(f_t;\lambda)}) \\ \frac{1}{2}(\sqrt{1+\rho(f_t;\lambda)} - \sqrt{1-\rho(f_t;\lambda)}) & \frac{1}{2}(\sqrt{1+\rho(f_t;\lambda)} + \sqrt{1-\rho(f_t;\lambda)}) \end{pmatrix},$$

$$\nabla_{\rho,t} = \frac{1}{1-\rho_t^2}\frac{1}{\nu-2+\epsilon_t^2+\eta_t^2}\left\{\nu\left[\epsilon_t\eta_t - \frac{1}{2}\rho_t(\epsilon_t^2+\eta_t^2-2)\right] + 2\left[\epsilon_t\eta_t - \rho_t\right]\right\}.$$

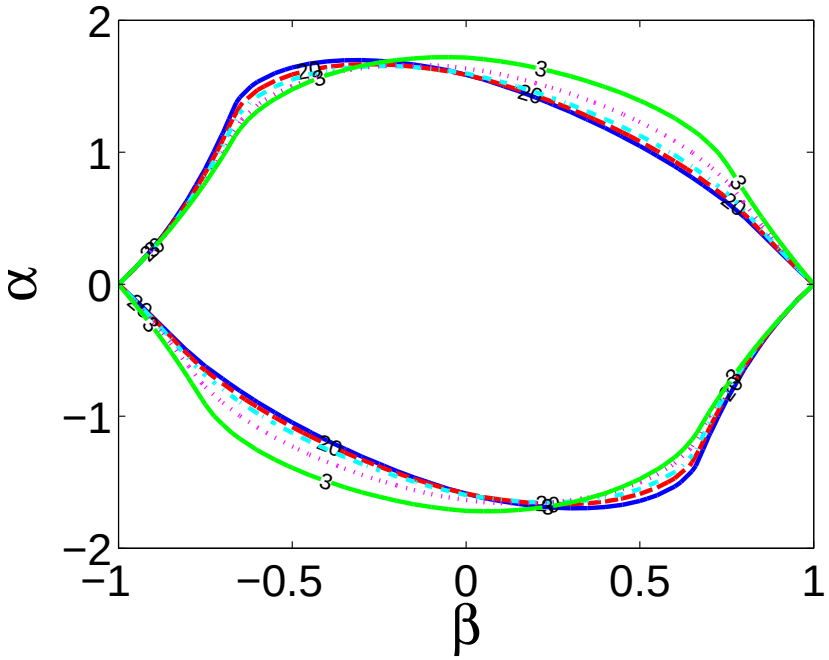
Notice from the above how the limiting case $\nu \to \infty$ reduces to updating corresponding to the normal distribution.

Appropriate scalings and transformations of the above then yield closed-form expressions for the updating equation. Also note that reparametrizing the correlation parameter by the Fisher transformation $\rho_t = \tanh(f_t)$, result in multiplying the score by a factor $(1-\rho_t^2)$, and thus the information matrix by a factor $(1-\rho_t^2)^2$ by an application of the chain rule.

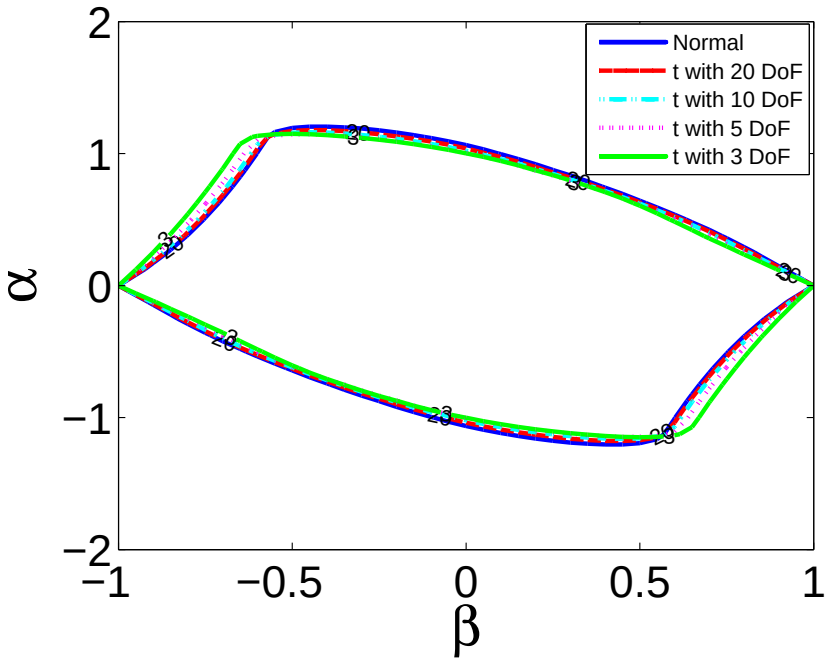
2.C Further numerical results

Figure 2.4 plots the SE regions for the Student's t case with different degrees of freedom. For the symmetric root case (panel (a)), in the relevant first quadrant lower degrees of freedom result in larger regions. The opposite holds for the Cholesky decomposition; see panel (b).

Figure 2.5 plots the results for $\psi(\rho) = k_\psi \arcsin(\rho)$. The left panel gives the result for the symmetric matrix root $k_\psi = 1/2$. The right panel is for the Cholesky decomposition, $k_\psi = 1$. Each panel presents 5 different regions. The outer region is based on the numerical evaluation of the original condition (2.8), with the infimum over ψ replaced by the choice $\psi(\rho) = \arcsin(\rho)/2$. The next region is obtained a numerical evaluation of (2.8) after applying Jensen's inequality, interchanging the expectations and the log operator. The next region follows after applying the triangle inequality, see the second line of equation (2.12). The final two regions are obtained after applying the Cauchy-Schwarz, or a second triangle inequality; see equations (2.12) and (2.13).

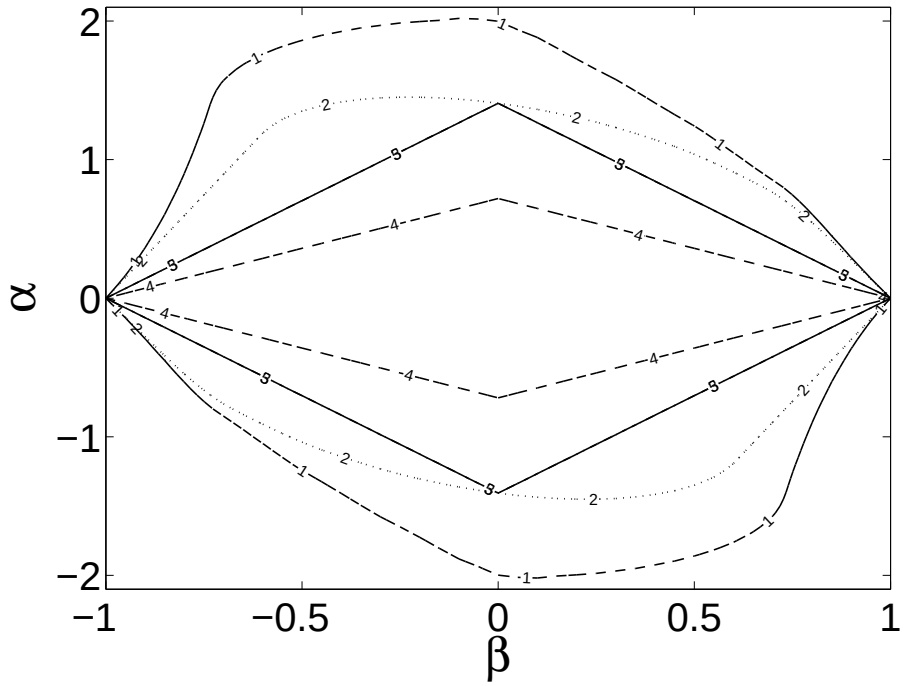


(a) Symmetric root ($k_\psi = 1/2$)

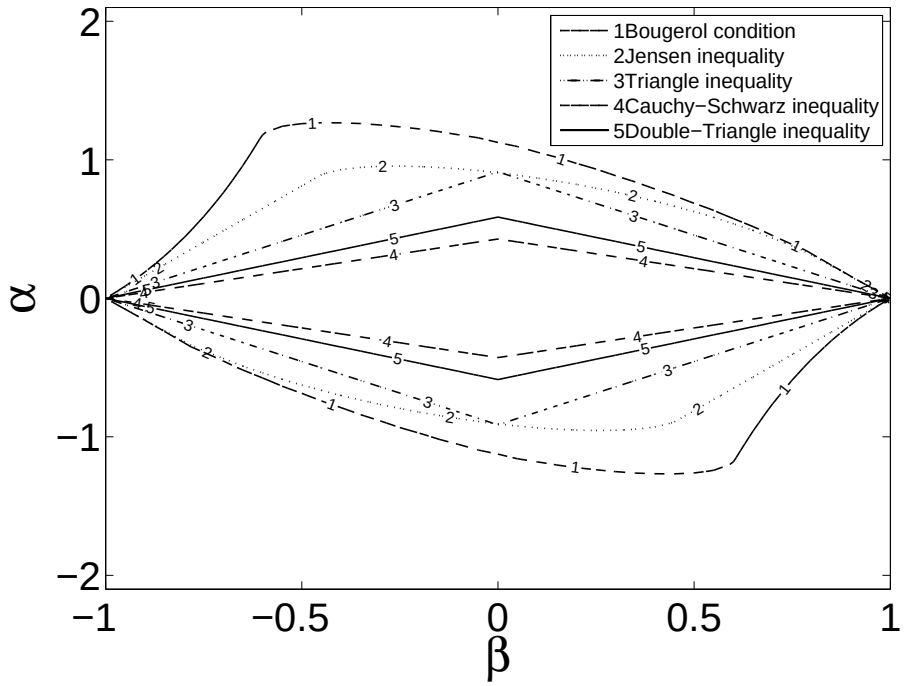


(b) Cholesky root ($k_\psi = 1$)

Figure 2.4: Contraction sufficiency regions for different Student's t degrees of freedom (DoF) and $a = 1/2$



(a) Symmetric root ($k_\psi = 1/2$)



(b) Cholesky root ($k_\psi = 1$)

Figure 2.5: Contraction sufficiency regions for the normal distribution using unit scaling ($S(f_t; \lambda) \equiv 1$) and the stricter inequalities in equation (2.12).

2.D Complementary results to the empirical application

2.D.1 FTSE: Diagnostic checks on the univariate models in the GJR family

First, we use the GJR family

$$\sigma^2(t) = c + g_1\sigma^2(t-1) + \ldots + g_P\sigma^2(t-P) + a_1 y(t-1)^2 + \ldots + a_Q y(t-Q)^2$$
$$+ L_1\, 1[y(t-1) < 0] y(t-1)^2 + \ldots + L_Q\, 1[y(t-Q) < 0] y(t-q)^2$$

The GJR(1,2) specification is preferred as it gives the log-likelihood of -5121.31 as opposed to -5127 of the GJR(1,1) model and -5129 of the EGARCH(1,2) model. Adding more lags is neither significant, nor do the diagnostics change.

Table 2.2: GJR(1,2) Conditional Variance Model for FTSE 100 Index

Parameter	Value	Standard Error	t- Statistic
Constant	0.0372	0.0089	4.1844
GARCH1	0.8741	0.0129	67.4199
ARCH1	0.0022	0.0132	0.1698
ARCH2	0.0755	0.0188	4.0119
Leverage1	0.0816	0.0189	4.3201
DoF	7.1869	0.9560	7.5173

Table 2.3: Results of Ljung-Box Test for Remaining ARCH effects (H0:none)

lags	Result Hypothesis	p-Value	Statistic	Crit Val
10	0	0.7749	6.4643	18.3070
15	0	0.8144	10.0840	24.9958
20	0	0.8900	12.6963	31.4104
25	0	0.8866	16.8668	37.6525
30	0	0.7492	24.4944	43.7730

Table 2.4: Results of Engle's Test for Remaining ARCH effects (H0:none)

lags	Result Hypothesis	p-Value	Statistic	Crit Val
10	0	0.7826	6.3780	18.3070
15	0	0.8018	10.2799	24.9958
20	0	0.8865	12.7830	31.4104
25	0	0.8919	16.7129	37.6525
30	0	0.7266	24.9676	43.7730

2.D.2 Athex Composite Index: Diagnostic checks on the univariate models in the EGARCH family

Here the EGARCH family

$$\begin{aligned}\log[\sigma^2(t)] = c &+ g_1 \log[\sigma^2(t-1)] + \ldots + g_P \log[\sigma^2(t-P)] \\ &+ a_1(|y(t-1)| - E|y(t-1)|) + a_Q(|y(t-Q)| - E|y(t-Q)|) \\ &+ L_1 y(t-1) + \ldots + L_Q y(t-Q)\end{aligned}$$

is chosen. EGARCH(1,2) is preferred as it gives a log-likelihood of -4041.02 as opposed to -4046.13 of the EGARCH(1,1) model and -4052.14 of the GJR(1,2) model. Adding more lags is neither significant, nor do the diagnostics change.

Table 2.5: EGARCH(1,2) Conditional Variance Model for Athex Composite Index

Parameter	Value	Standard Error	t- Statistic
Constant	-0.0001	0.0019	-0.0548
GARCH1	0.9824	0.0024	407.9410
ARCH1	-0.0192	0.0413	-0.4649
ARCH2	0.1434	0.0429	3.3374
Leverage1	-0.1304	0.0108	-12.0011
DoF	12.8521	2.8669	4.4828

Table 2.6: Results of Ljung-Box Test for Remaining ARCH effects (H0:none)

lags	Result Hypothesis	p-Value	Statistic	Crit Val
10	0	0.8610	5.4257	18.3070
15	0	0.5087	14.2230	24.9958
20	0	0.5506	18.5588	31.4104
25	1	0.0261	40.4590	37.6525
30	1	0.0399	44.8484	43.7730

2.D.3 Robustness check: Modeling both marginals by an EGARCH(1,2) model

The results in this section show that specification changes in the volatility equations leave our results for correlation models practically unaltered. We model both series by an EGARCH(1,2). We omit the robustness check of modeling both series by

Table 2.7: Results of Engle's Test for Remaining ARCH effects (H0:none)

lags	Result Hypothesis	p-Value	Statistic	Crit Val
10	0	0.8653	5.3680	18.3070
15	0	0.5097	14.2092	24.9958
20	0	0.5547	18.4975	31.4104
25	1	0.0390	38.7546	37.6525
30	0	0.0546	43.3479	43.7730

as for the Athex Composite Index, as the GJR binding positivity result in an inferior model fit (a likelihood decrease of more than 10 points).

Table 2.8: Both EGARCH(1,2) marginals. Full Estimation Results.

Notes: Heteroskedasticity and autocorrelation consistent (HAC) standard errors in parentheses. Best log-likelihood, AIC, and BIC values across models are printed in boldface. $H_0^{5\%}$: $\rho_t \equiv 0$ and $H_0^{5\%}$: $\rho_t \equiv \bar{\rho}$ indicate whether the Nyblom test among residuals rejects constant zero or estimated correlation, respectively. HAC standard errors are computed using Newey-West weights with $\min(\lfloor 1.2 \times T^{1/3} \rfloor, T)$ lags.

	EWMA1	EWMA2	EWMA3	t(∞)GAS (a=0)	t(∞)GAS (a=0.5)	t(∞)GAS (a=1)	t(5)GAS (a=0)	t(5)GAS (a=0.5)	t(5)GAS (a=1)	t($\cdot$)GAS (a=1)
λ	5	5	8.6648	∞	∞	∞	5	5	5	8.9598
			(0.00742)							(0.8788)
c		0.0002	0.0010	0.0115	0.0112	0.0109	0.0089	0.0089	0.0088	0.0112
		(0.0003)	(0.0003)	(0.0448)	(0.0451)	(0.0358)	(0.0111)	(0.0113)	(0.0067)	(0.0071)
A				0.0254	0.0281	0.0310	0.0334	0.0315	0.0296	0.0356
				(0.0481)	(0.0472)	(0.0416)	(0.0172)	(0.0144)	(0.0089)	(0.0100)
B	0.9771	0.9769	0.9761	0.9757	0.9763	0.9770	0.9798	0.9799	0.9801	0.9766
	(0.0069)	(0.0048)	(0.0019)	(0.1098)	(0.0914)	(0.0723)	(0.0237)	(0.0240)	(0.0142)	(0.0141)
Log-likelihood In-sample	−7845	−7844	−7812	−7886	−7886	−7886	−7838	−7838	−7838	−7803
AIC	15691	15692	15629	15779	15779	15778	15682	15682	15682	15615
BIC	15697	15704	15647	15797	15797	15796	15700	15700	15700	15639
# estimated parameters	1	2	3	3	3	3	3	3	3	4
$H_0^{5\%}$: $\rho_t \equiv 0$	reject	reject								
$H_0^{5\%}$: $\rho_t \equiv \bar{\rho}$	reject	reject	reject	reject	reject		reject	reject	reject	
Inside SE region?										
$k_\psi = 1$ (Cholesky)				Yes	Yes	No	No	No	No	No
$k_\psi = 1/2$ (Symmetric)				Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The second line underneath each parameter estimate depicts HAC sandwich standard errors. For volatility, EGARCH(1,2) marginals are used for both series.

Chapter 3

High-Dimensional Dynamic Correlation Models: Formulations, Properties and Estimation

Abstract. We highlight the similarities and the differences between the updating mechanisms of the dynamic conditional correlation (DCC) based equicorrelation model of Engle and Kelly (2012) and the recent score-based dynamic correlation models of Creal, Koopman, and Lucas (2011, 2013). For score driven models, we provide new theoretical properties by establishing conditions for stationarity and ergodicity of the model and conditions for consistency of the quasi maximum likelihood estimator. In an empirical application, we propose a joint model for both equity returns and options data to measure time-varying correlation risk premia.

3.1 Introduction

The estimation of dynamic dependence measures for asset returns, interest rates and credit spreads is an important research topic in financial econometrics and risk management. A particular challenge in multivariate dependence modeling is the rapid increase in the number of both dynamic and static model parameters if the dimension of the data increases. This has led many researchers to propose models that combine parsimony with a sufficient degree of flexibility to accommodate multivariate volatility and correlation dynamics; see for example Bauwens, Laurent, and Rombouts (2006) for an overview.

One of the most striking findings in the recent literature is that it is often hard for models with a complex dynamic correlation structure to beat models with an extremely simple correlation structure, particularly in high-dimensional settings. This holds even more if one accounts for the potential effect of estimation uncertainty. A key reference in this respect is the dynamic equicorrelation (DECO) model of Engle and Kelly (2012). They use an extremely simple correlation matrix structure defined by one (or at most a few) dynamic parameters and show that such a model does well in forecasting and assessing portfolio risk dynamics. Engle and Kelly show in a controlled simulation setting that their equicorrelation model outperforms a correctly specified but estimated dynamic conditional correlation (DCC) model of Engle (2002a). These findings are in line with results from the finance literature showing that simple asset allocation strategies that spread out wealth evenly

across all available assets (and thus implicitly assume an equicorrelation structure) have a strong out-of-sample performance. The strong performance of these naive asset allocation strategies is often superior to that of optimal mean-variance portfolios where all means, variances, and covariances are estimated simultaneously; see DeMiguel, Garlappi, and Uppal (2009) and Vasnev, Claeskens, Magnus, and Wang (2014).

Recently, alternative versions of the equicorrelation model have been proposed in the literature. Apart from the original formulation by Engle and Kelly (2012) that uses the dynamic conditional correlation (DCC) dynamics of Engle (2002b), a number of robust, score-based models for dynamic correlations and copulas have been proposed. Examples include Creal, Koopman, and Lucas (2011, 2013), Harvey (2013), Oh and Patton (2013), Lucas et al. (2013, 2014), De Lira Salvatierra and Patton (2013), Bernardi and Catania (2015), and Avdulaj and Barunik (2015). So far, however, little attention has been paid to how these different equicorrelation models are related. For example, it is not clear how different specifications exploit and combine the information in the data differently, and how these different ways can be more or less efficient in specific settings. Second, little is known about the statistical theory of the new score-driven (equi)correlation models. In particular, we do not know under what conditions score-driven equicorrelation models are stationary and ergodic. This is the main theoretical contribution of the current paper.

First, we show that score-driven equicorrelation models and the DCC equicorre-

lation model handle the contemporary and dynamic effects of influential observations in a different way. To make a consistent comparison of filtered correlation paths (in the sense of Patton, 2011), we use a quasi maximum likelihood (QML) framework to estimate the models' static parameters. Such an approach is new to the generalized autoregressive score (GAS) literature, which so far mainly focuses on full maximum likelihood (ML) estimation. The advantage of QML over ML is that we mitigate part of the sensitivity of ML when using a possibly misspecified likelihood, while at the same time retaining a number of the attractive robustness features of fat-tailed score-driven volatility models; see the discussions in Boudt et al. (2011), Creal et al. (2011, 2013) and Harvey (2013).

The second contribution of the current paper is to characterize the stochastic properties of score-driven correlation models. We formulate new conditions for stationarity and ergodicity that substantially simplify and improve upon the results in Blasques et al. (2016). Furthermore, the new results are not only applicable to correlation models, but more generally to multivariate scale models. Dynamic multivariate models necessarily introduce non-linearities into the specification that are challenging for establishing stationarity and ergodicity. For example, only recently a sufficient condition for strict stationarity has been given by an analysis of algebraic geometry for general BEKK processes; see Boussama et al. (2011). However, as we will argue, a surprisingly simple pattern emerges in a subclass of score-driven correlation models, where the nonlinearity enforces stability.

To illustrate the usefulness of our results, we apply the equicorrelation model to a U.S. data set of stock returns and option implied correlations. We use the equicorrelation model to estimate the dynamic correlation structure of U.S. stock returns. In this setting, QML estimation is advantageous as it is economically interpretable, while the implications of a full density fit for a possibly severely mis-specified model may be questionable. Furthermore, a QML based estimation procedure of the score-driven model is numerically efficient in such high-dimensional settings, as for the Gaussian observation density the estimation problem can be split into two separate stages: in a first step volatilities are estimated and in a second step one applies the estimated volatilities to extract dynamic correlations. We combine the dynamic equicorrelation model with two option implied correlation measures obtained from the CBOE into a new joint model. This allows us to investigate whether correlations bear a time-varying risk premium. The importance of (static) correlation risk premia is well-known from the empirical finance literature; see for instance Pollet and Wilson (2010). Using data over the period 2002–2013, we find that the risk premia are much more persistent than the equity induced correlation measures. This difference in persistence can help to explain why risk-neutral correlations have been found to be bad predictors of their physical counterparts, in contrast to the results for risk-neutral volatilities.

The remainder of this paper is organized as follows. Section 3.2 introduces the generalized autoregressive score equicorrelation model and discusses its relation to

the dynamics embedded in the DCC equicorrelation model of Engle and Kelly (2012). Section 3.3 derives the asymptotic statistical properties of the QMLE for score-driven equicorrelation models. Section 3.5 presents our empirical application using equity return data and option implied correlations. Section 3.6 concludes. The Appendix gathers the proofs.

3.2 Dynamic equicorrelation models

3.2.1 The model and its quasi-likelihood function

Consider the equicorrelation model

$$y_t = \sqrt{\rho_t}\,\iota_n\, g_t + \sqrt{1-\rho_t}\,\varepsilon_t, \quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \mathrm{I}_n)\,, \tag{3.1}$$

where $y_t = (y_{1t}, \ldots, y_{nt})^\top \in \mathbb{R}^n$ for $t = 1, \ldots, T$, denotes the vector of observations, the correlation parameter ρ_t with values in $(-(n-1)^{-1}, 1)$ determines the relative magnitude of the i.i.d. common factor g_t with zero mean and unit variance, and the idiosyncratic components $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})^\top$. We assume g_t and ε_t are independent. It follows that the covariance matrix of y_t is given by

$$\Sigma_t = \rho_t \iota_n \iota_n^\top + (1-\rho_t)\mathrm{I}_n.$$

If y_t has an elliptical distribution, computing the likelihood function only requires the computation of the log determinant and the inverse of Σ_t. These two quantities are given analytically by the expressions

$$\Sigma_t^{-1} = \frac{1}{1-\rho_t}\mathrm{I}_n - \frac{\rho_t}{(1-\rho_t)\,(1+(n-1)\rho_t)}\iota_n\iota_n^\top, \tag{3.2}$$

$$\det(\Sigma_t) = (1-\rho_t)^{n-1}\,(1+(n-1)\rho_t)\,. \tag{3.3}$$

Note that all elements of y_t in our specification have unit variance. We specify the model as such to fully concentrate on the correlation dynamics. The covariance matrix Σ_t in our setting can thus be viewed as a correlation matrix, and we indicate this from now on by writing

$$\Sigma_t = R^{\mathrm{DECO}}(\rho_t) = \begin{pmatrix} 1 & \rho_t & \dots & \rho_t \\ \rho_t & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_t \\ \rho_t & \dots & \rho_t & 1 \end{pmatrix} = \rho_t\iota_n\iota_n^\top + \big(1-\rho_t\big)\mathrm{I}_n. \tag{3.4}$$

If variances are also stochastic, we can easily transform the model to measure $\tilde{y}_t = D_t^{-1}y_t$, with a diagonal matrix D_t containing the volatilities of y_{it} for $i = 1,\dots,n$ and $t = 1,\dots,T$.

In this paper, we follow Engle (2002a) and adopt a Gaussian quasi-likelihood for y_t to estimate the model's parameters. Using the Gaussian quasi-likelihood as our

criterion function, it is well-known that we can separate the estimation of volatilities and correlations into two steps. This is a desirable feature in high-dimensional systems such as ours due to the computational gains involved. Moreover, estimation of the model's static parameters is still consistent, even if the Gaussian quasi-likelihood is mis-specifcied. Using equations (3.2) and (3.3), the Gaussian quasi-likelihood can be written as $\mathcal{L}_T = \sum_{t=1}^{T} \ell_t$, with,

$$\ell_t = -\tfrac{1}{2}\left(\log\left((1-\rho_t)^{n-1}(1+(n-1)\rho_t)\right) + \frac{y_t^\top y_t}{1-\rho_t} - \frac{\rho_t \cdot \left(\iota^\top y_t\right)^2}{(1-\rho_t)(1+(n-1)\rho_t)}\right). \tag{3.5}$$

The quasi-maximum likelihood (QML) criterion function only depend on the quantities $\iota^\top y_t$ and $y_t^\top y_t$, which hold the sum and sum of squares of the elements in y_t, respectively. Therefore, these are the two sufficient statistics of the model.[1] With these two statistics and the value of ρ_t, the QML criterion can be computed very efficiently, even in high-dimensional settings. The efficiency of the entire estimation procedure then hinges on the dynamic model for the correlation parameter ρ_t.

[1]Recall that a statistic $T(y)$ is sufficient for the parameter f if the conditional distribution of y given $T(y)$ does not depend on f. By the Fisher- Neyman factorization theorem, we can then equivalently rewrite the likelihood as

$$p_y(y|f) = h(y)\, g_\theta(T(y)|f)$$

see also (3.5) for our concrete equicorrelation example.

3.2.2 DCC-DECO dynamics

Engle and Kelly (2012) specify a dynamic model for ρ_t based on the Dynamic Conditional Correlation (DCC) approach[2] of Engle (2002a) and Aielli (2013). We indicate this by writing $\rho_t = \rho_t^{\text{DCC}}$. We have

$$Q_{t+1} = (1 - \alpha - \beta)\bar{Q} + \beta Q_t + \alpha \tilde{Q}_t^{1/2} y_t y_t^\top \tilde{Q}_t^{1/2}, \tag{3.6}$$

$$R_t^{\text{DCC}} = \tilde{Q}_t^{-1/2} Q_t \tilde{Q}_t^{-1/2}, \qquad \rho_t^{\text{DCC}} = \frac{1}{n(n-1)} \iota^\top \left(R_t^{\text{DCC}} - \mathrm{I}_n \right) \iota, \tag{3.7}$$

where $\alpha, \beta > 0$ are scalars, $\tilde{Q}$ replaces the off-diagonal elements of Q with zeros, and the targeting matrix $\bar{Q}$ is obtained from the unconditional correlation matrix of the standardized observations $y_t^* = \tilde{Q}_t^{1/2} y_t$; see also Aielli (2013). Note that even though ρ_t^{DCC} is univariate, we still need to keep track of the $n \times n$ matrices Q_t and R_t^{DCC} for all $t = 1, \ldots, T$. Moreover, we need to estimate the entire targeting matrix $\bar{Q}$. We come back to both of these issues later.

3.2.3 GAS-DECO dynamics

Recently, a number of papers have proposed alternative updating mechanisms for the correlation parameters based on the score of the predictive likelihood. These models are known as generalized autoregressive score (GAS) models; see Creal et al. (2013) and Harvey (2013). Correlation and copula versions of these models have been

[2]Even though we use the terms DCC and cDCC interchangeably throughout, we have implemented the cDCC of Aielli (2013) in all computations.

considered in for example Creal et al. (2011), Boudt et al. (2012), Lucas et al. (2014), Bernardi and Catania (2015), and De Lira Salvatierra and Patton (2013).

Consider the multivariate scale model

$$y_t = h(f_t)\, u_t, \qquad u_t \overset{\text{i.i.d.}}{\sim} (0, \mathrm{I}_n), \qquad t \in \mathbb{N} \tag{3.8}$$

$$f_{t+1} = \omega + \beta\, f_t + \alpha\; s\big(f_t, y_t; \lambda\big), \tag{3.9}$$

$$s_t := s\big(f_t, y_t; \lambda\big) = S\big(f_t; \lambda\big) \cdot \nabla\big(f_t, y_t; \lambda\big) \tag{3.10}$$

$$\nabla_t := \nabla\big(f_t, y_t; \lambda\big) = \partial \log p_y\big(y_t | f; \lambda\big)/\partial f \,\big|_{f=f_t}, \tag{3.11}$$

initialized at some fixed initial value $f_1 \in \mathbb{R}$, where $h(f_t)$ is a scale matrix; u_t has a density function $p_u(u_t; \lambda)$ with zero mean, unit covariance matrix I_n, and depending on the parameter vector λ, such as the degrees of freedom parameter if p_u is a Student's t density; ω, α and β are elements of the static parameter vector $\theta \in \Theta$; $h : \mathbb{R}^k \to \mathbb{R}^{n \times n}$ maps the time-varying parameter vector f_t to the positive definite matrix $h(f_t)$; $S(f_t; \lambda) \in \mathbb{R}^{k \times k}$ is a scaling matrix for the score $\nabla(f_t, y_t; \lambda)$ function of the conditional density $p_y(y_t | f_t; \lambda)$ of the observations (3.8).[3] It is clear that (3.8) encompasses (3.1) by defining $h(f_t)$ such that $h(f_t)h(f_t)^\top = \Sigma_t$. Note that the distribution p_u of u_t

[3]The model in (3.8) covers a large number of time-varying volatility and correlation models by an appropriate choice for the dynamics of f_t. Examples include the BEKK model of Engle and Kroner (1995a), the CCC model of Bollerslev (1990), the DCC model of Engle (2002a) and its modification by Aielli (2013), the OGARCH model of Ding (1994), the GOGARCH model of Van der Weide (2002); see also Bauwens et al. (2006) for an overview. Moreover, (3.8) also contains the recent score-driven multivariate volatility and correlation models of Creal et al. (2011) and Lucas et al. (2014), and the univariate Beta-t-GARCH model of Harvey and Chakravarty (2008) and Harvey (2013), which are all special cases of the Generalized Autoregressive Score modeling framework of Creal et al. (2013).

may depend on the parameter λ, for example the degrees of freedom parameter of a Student's t distribution.

The following two examples illustrate the properties of score-driven correlation models in two familiar cases and prelude to the comparative analysis with the DCC equicorrelation model of Engle and Kelly (2012) in Section 3.2.4.

Example 1 (Bivariate Gaussian)**.** For a bivariate normal distribution with unit variances and link function $\rho_t = \tanh(f_t)$, the (unscaled) score $s(\rho_t, y_t; \lambda)$ of the normal distribution with respect to ρ_t is given by

$$s(\rho_t, y_t; \lambda) = \frac{1}{1-\rho_t^2} \cdot \Big[(1+\rho_t^2)(y_{1t}y_{2t} - \rho_t) - \rho_t(y_{1t}^2 + y_{2t}^2 - 2)\Big] \tag{3.12}$$

$$= \frac{1}{1-\rho_t^2} \cdot \Big[\tfrac{1}{2}(1+\rho_t^2)(y_{1t}+y_{2t})^2 - \tfrac{1}{2}(1+\rho_t)^2(y_{1t}^2+y_{2t}^2) + \rho_t(1-\rho_t^2)\Big] \tag{3.13}$$

$$= \frac{1}{1-\rho_t^2} \cdot \Big[(1-\rho_t)^2(y_{1t}y_{2t} - \rho_t) - \rho_t(y_{1t}-y_{2t})^2 + 2\rho_t(1-\rho_t)\Big]. \tag{3.14}$$

This score has a clear interpretation. The form in equation (3.12) reveals that the score is a martingale difference: it has zero conditional mean if the model is well-specified as $\mathbb{E}[y_{1t}y_{2t}] = \rho_t$ and $\mathbb{E}[y_{1t}^2 + y_{2t}^2] = 2$. We exploit this feature later on to establish stationarity and ergodicity properties of the correlation process f_t in (3.9). The specification in equation (3.13) is also interesting in that it shows that the score is essentially formulated in terms of two sufficient statistics, namely $(y_{1t} + y_{2t})$, and $(y_{1t}^2 + y_{2t}^2)$. This feature can be extended more generally to the equicorrelation model.

We can exploit this feature to operationalize the model in high-dimensional settings. Note that these are precisely the same two sufficient statistics as needed to evaluate the QML criterion (3.5). Finally, the formulation in (3.14) allows us to relate the score-driven approach to alternative models for time varying correlations. As we show later, the first term $(y_{1t}y_{2t} - \rho_t)$ in (3.14) is similar to the DCC updating scheme of Engle (2002a). The second term on the other hand, $(y_{1t}-y_{2t})^2$, captures the differences between observations and thus between their order statistics. As such, it resembles the specification of Patton (2006b) for general copulas with time varying parameters. The approach based on the score combines these two sources of information. When the correlation increases, deviations in expected tail dependence $(y_{1t} - y_{2t})^2$ become more dominant in the updating recursion. We provide more details on the comparison between the score-driven and DCC approach in Section 3.2.4.

We also note that we can write the score in (3.14) in terms of two independent standard normal random variables u_{1t} and u_{2t} by defining $y_{1t} = u_{1t}$ and $y_{2t} = \rho_t u_{1t} + \sqrt{1-\rho_t^2}\, u_{2t}$. We obtain

$$s_u(\rho_t, u_t; \lambda) = s(\rho_t, R^{\text{DECO}}(\rho_t)^{1/2} u_t; \lambda) = \sqrt{1-\rho_t^2}\, u_{1t}u_{2t} - \rho_t(u_{2t}^2 - 1). \tag{3.15}$$

The above expression illustrates that the trade-off between concordance $(u_{1t}u_{2t})$ and magnitude of independent innovation pairs $(u_{2t}^2 - 1)$ takes place on the unit circle with weights $(1-\rho_t^2)^{1/2}$ and ρ_t, respectively. We make extensive use of this feature

of bounded weights in developing the more general statistical theory for the equicorrelation model later on.

Example 2 (Bivariate Student's t)**.** For fat-tailed densities, the GAS framework suggests a robust filter for variances and correlations. Take $f_t = \rho_t$ and let $s(\rho_t, y_t; \lambda)$ be the score of a bivariate Student's t distribution with unit variances and correlation parameter ρ_t. After some algebraic manipulations it follows directly from Creal et al. (2011) that

$$s(\rho_t, y_t; \lambda) = \frac{1}{1-\rho_t^2} \cdot \Big[\tfrac{1}{2} w_t (1+\rho_t^2)(y_{1t}+y_{2t})^2 - \tfrac{1}{2} w_t (1+\rho_t)^2 (y_{1t}^2 + y_{2t}^2) + \rho_t (1-\rho_t^2) \Big], \tag{3.16}$$

$$w_t = \big(\lambda + n\big) \Big/ \left(\lambda - 2 + \frac{y_{1t}^2 - 2\rho_t y_{1t} y_{2t} + y_{2t}^2}{1-\rho_t^2} \right),$$

where $\lambda > 2$ denotes the degrees of freedom parameter of the Student's t distribution with unit variance. The weight w_t decreases if y_t lies more in the tails of the distribution, such that tail observations have a smaller impact on volatilities and correlations compared to the Gaussian case. This provides the robustness features to the score-driven Student's t based correlation model, as mentioned earlier. We note again that the score can be formulated in terms of the two sufficient statistics $(\iota^\top y_t)^2$ and $y_t^\top y_t$, a feature that extends to the entire class of multivariate scale models for with an equicorrelation structure and elliptically distributed u_t; see Section 3.3.

The specification in (3.9) only has one lag of $s(f_t, y_t; \lambda)$ and f_t. It is straight-

forward, however, to include more lags, non-linearities, or even long-memory features into (3.9); see for example Creal et al. (2013), Harvey and Luati (2014), and Janus et al. (2014). The use of the transformed Cholesky decomposition $h(f_t)$ in equation (3.8) moreover allows us to reparametrize the model such that the correlation ρ_t automatically lies in the correct range. For example, in the bivariate case, we can use the transformation $\rho_t = \tanh(f_t)$, such that $-1 < \rho_t < 1$ for any value of $f_t \in \mathbb{R}$; see also Van der Vaart (2000, Example 3.6, p. 30). Choosing an appropriate transformation in the multivariate context can be a non-trivial issue; see also Laurent et al. (2009). In our general n-dimensional equicorrelation setting, we use the adjusted tanh parameterization

$$\rho_t^{\text{GAS}} = \rho(f_t) = \tanh_n^*(f_t) = \tfrac{1}{2}\left(1 - \frac{1}{n-1}\right) + \tfrac{1}{2}\left(1 + \frac{1}{n-1}\right)\tanh(f_t), \qquad (3.17)$$

which forces $\rho_t(f_t) \in (-1/(n-1), 1)$ for every $f_t \in \mathbb{R}$ and every dimension n. The score-driven approach automatrically take these reparametrizations into account for the dynamics of f_t. For example, in the bivariate normal example the tanh reparametrization requires us to postmultiply the original score with respect to ρ_t in (3.14) by the derivative $\dot{\rho}(f_t) = \partial\,\rho(f)/\partial f \mid_{f=f_t}$ due to a chain rule argument. In case of the tanh parametrization, this amounts to multiplying (3.14) by the factor $\dot{\rho}(f_t) = (1 - \rho(f_t)^2)$, which cancels the effect of the denominator in (3.14). A clever reparameterization can thus remove possible degeneracies from the score, such as pos-

sible division by zero if the correlation tends to 1. The $\tanh_n^*$ parameterization in (3.17) accomplishes the same in a general dimension n. The $\tanh_n^*$-reparametrization also has the advantage that predetermined regressors can be easily be included, similar to GARCH-X models. Due to the link function in (3.17), interactions with economic variables and average market correlations can be studied. This also motivates our joint model for equity returns and option-implied correlations in Section 3.5.

The GAS dynamics in (3.9) use the scaled score $s(f_t, y_t; \lambda)$ at time t to improve the local fit of the model as measured by the log density $\log p_y(y_t|f_t; \theta)$. The model has information theoretical optimality properties in that it locally improves the Kullback-Leibler divergence upon each iteration step, even in cases where the statistical model is severely mis-specified; see Blasques et al. (2015). The model is observation-driven as defined by Cox (1981). For the scaling function $S(f_t; \lambda)$ of the score, Creal et al. (2013) advocate the use of powers of the conditional Fisher information matrix,

$$S(f_t; \lambda) = \mathcal{I}(f_t; \lambda)^{-a} \qquad \mathcal{I}_t(f_t; \lambda) = \mathbb{E}_{t-1}\left[\nabla_t(f_t, y_t; \lambda)\nabla_t(f_t, y_t; \lambda)^\top\right], \tag{3.18}$$

where a is typically taken as 0, 1/2 or 1. This accounts for the local expected curvature of the likelihood.

The following general result will prove useful.

Proposition 1. *The score expressions for the GAS model in equations* (3.8)–(3.11)

satisfies

$$\nabla_{u,t} = \nabla(f_t, h(f_t)u_t) = \Psi(f_t;\lambda)^\top \operatorname{vec}\left(-\nabla_{p_{u,\lambda}}(u_t)u_t^\top - \mathrm{I}_n\right), \tag{3.19}$$

$$\nabla_t = \nabla(f_t, y_t;\lambda) = \Psi(f_t;\lambda)^\top \operatorname{vec}\left(-\nabla_{p_{u,\lambda}}\left(h(f_t)^{-1}y_t;\lambda\right)\left(h(f_t)^{-1}y_t\right)^\top - \mathrm{I}_n\right), \tag{3.20}$$

$$\Psi_t = \Psi(f_t;\lambda) = \left(\mathrm{I}_n \otimes h(f_t;\lambda)^{-1}\right)\partial \mathrm{vec}(h(f_t;\lambda))/\partial f_t^\top, \tag{3.21}$$

where $\nabla_{p_{u,\lambda}}(u_t) = \nabla_{p_{u,\lambda}}(u_t;\lambda) = \partial \log p_u(u_t;\lambda)/\partial u_t$ *does not depend on the time-varying parameter* f_t*. The information matrix takes the form*

$$\mathcal{I}_t(f_t;\lambda) = \mathbb{E}\left[\nabla(f_t,u_t;\lambda)\nabla(f_t,u_t;\lambda)^\top\right] = \Psi_t^\top\left(\mathcal{I}_{p_{u,\lambda}} - \mathrm{vec}(\mathrm{I}_n)\mathrm{vec}(\mathrm{I}_n)^\top\right)\Psi_t, \tag{3.22}$$

$$\mathcal{I}_{p_{u,\lambda}} = \mathbb{E}[u_t u_t^\top \otimes \nabla_{p_{u,\lambda}}(u_t)\nabla_{p_{u,\lambda}}(u_t)^\top]. \tag{3.23}$$

An interesting feature about the score in (3.19) is that the parameterization $h(f_t;\lambda)$ and the shape of the error density p_u enter in a clear and almost separable way. The error distribution p_u enters through the score $\nabla_{p_{u,\lambda}}(u_t)$, while the choice of parameterization $h(f_t;\lambda)$ enters via the matrix $\Psi(f_t;\lambda)$. Such a property turns out to be useful for deriving some of the asymptotic properties of y_t as generated by (3.8)–(3.11).

Using Proposition 1, we can derive the expressions for the score of the equicorrelation model. If we moreover assume that the innovations u_t are elliptically distributed, the relevant scores take a simple and intuitive form. We summarize this in the fol-

lowing corollary. In this corollary, we use the generalized tanh parameterization in (3.17).

Corollary 1. *Consider the equicorrelation model with parameterization* (3.17) *and with elliptically distributed error terms* u_t *such that* $\nabla_{p_{u,\lambda}}(u_t) = -w_{u,t} \cdot u_t$ *with weights* $w_{u,t} := w_u(u_t^\top u_t; \lambda)$, *where* $w_u(\cdot; \lambda) : \mathbb{R}^+ \to \mathbb{R}^+$. *Define the functions* $\rho_{j,t} := \rho_j(\rho(f_t))$ *for* $j = 1, \ldots, 6$ *as in equations* (3.49)–(3.54) *of the Appendix. Then*

$$\nabla_u(f_t, u_t; \lambda) \;:=\; \rho_{1,t}\; w_{u,t} \cdot u_t^\top u_t + \rho_{2,t}\; w_{u,t} \cdot (u_t^\top \iota)^2 - (\rho_{1,t} + \rho_{2,t})\, n, \tag{3.24}$$

$$\nabla(f_t, y_t; \lambda) \;=\; \rho_{3,t}\; w_{y,t} \cdot y_t^\top y_t + \rho_{4,t}\; w_{y,t} \cdot (y_t^\top \iota)^2 - (\rho_{1,t} + \rho_{2,t})\, n, \tag{3.25}$$

where

$$w_{y,t} \;:=\; w_y(f_t, y_t; \lambda) := w_u\big(\rho_{5,t}\; y_t^\top y_t + \rho_{6,t}\; (y_t^\top \iota)^2; \lambda\big). \tag{3.26}$$

The weight $w_{y,t}$ in (3.26) accounts for possible fat-tailedness and other features of the density p_u. If p_u is Gaussian, we have $w_{y,t} \equiv 1$, while for example for the Student's t case with $\lambda > 2$ degrees of freedom and variance 1, we have

$$w_{u,t} = \frac{\lambda + n}{\lambda - 2 + u_t^\top u_t}, \tag{3.27}$$

resulting in the weight in (3.16) in Example 2. The Student's t weight for $\lambda < \infty$ thus decreases if $\|u_t\|$ is large, i.e., if the error lies far in the tails. This gives a robustness

feature to the dynamics of the GAS specification; see Creal et al. (2011) for more details.

Equations (3.25) and (3.26) directly reveal that the GAS-DECO model from Section 3.2 for elliptical distributions only requires us to keep track of two univariate sufficient statistics, namely $y_t^\top \iota$, and $y_t^\top y_t$. Unlike for example the DCC-DECO model, we no longer need a recursion in the $n \times n$ matrices Q_t. Given that the Gaussian QML criterion in (3.5) depends on the same two univariate sufficient statistics, the GAS-DECO model retains it numerical efficiency for any elliptical density p_u.

3.2.4 Comparison of GAS-DECO and DCC-DECO dynamics

It is interesting to highlight some of the differences between the original DECO dynamics of Engle and Kelly (2012) as specified in equations (3.6) and (3.7), and the dynamics of the score-driven equicorrelation model under fat tails. Using equations (3.6) and (3.7), we obtain

$$\rho_{t+1}^{\text{DCC}} = \frac{\iota^\top \left(\tilde{Q}_{t+1}^{-1/2} Q_{t+1} \tilde{Q}_{t+1}^{-1/2} - \mathrm{I}_n \right) \iota}{n(n-1)} = \tag{3.28}$$
$$\frac{\iota^\top \left[(1-\alpha-\beta) \left(\bar{q}_{t+1} \bar{R} \, \bar{q}_{t+1}^\top - \mathrm{I}_n \right) + \beta \left(\tilde{q}_{t+1} R_t^{\text{DCC}} \, \tilde{q}_{t+1}^\top - \mathrm{I}_n \right) + \alpha \left(\tilde{q}_{t+1} \, y_t y_t^\top \, \tilde{q}_{t+1}^\top - \mathrm{I}_n \right) \right] \iota}{n(n-1)},$$

with diagonal matrices $\bar{q}_t = \tilde{Q}_t^{-1/2}\tilde{\bar{Q}}^{1/2}$ and $\tilde{q}_t = \tilde{Q}_t^{-1/2}\tilde{Q}_{t-1}^{1/2}$, and correlation targeting matrix $\bar{R} = \tilde{\bar{Q}}^{-1/2}\bar{Q}\tilde{\bar{Q}}^{-1/2}$, with equicorrelation target $\bar{\rho} = \iota^\top(\bar{R} - \mathrm{I}_n)\iota/(n(n-1))$. To provide an intuition for the differences between GAS-DECO and DCC-DECO transition dynamics, we first consider the special case where the diagonal elements of $\tilde{Q}_t$ are 'in equilibrium', i.e., $\tilde{\bar{Q}} = \tilde{Q}_t = \tilde{Q}_{t-1}$ and $\bar{q}_t = \tilde{q}_t = \mathrm{I}_n$. Equation (3.28) then simplifies to

$$\rho_{t+1}^{\mathrm{DCC}} = (1 - \alpha - \beta)\,\bar{\rho} + \beta\,\rho_t^{\mathrm{DCC}} + \alpha\,\Delta_t^{\mathrm{DCC}}, \qquad \Delta_t^{\mathrm{DCC}} = \frac{(\iota^\top y_t)^2 - n}{n(n-1)}. \tag{3.29}$$

If the DCC model is correctly specified, we have the unbiasedness property $\mathbb{E}_{t-1}[\Delta_t^{\mathrm{DCC}}] = \rho_t^{\mathrm{DCC}}$ at the true parameters. The assumed condition for covariance stationarity is $|\alpha + \beta| < 1$.

As can be seen in Example 2, the dynamics of the GAS-DECO correlation parameter for Student's t distributed u_t take the form

$$\begin{aligned} \rho_{t+1}^{\mathrm{GAS}} &= \rho_{t+1}\left(\omega + \beta\,\rho_t^{\mathrm{GAS}} + \alpha\,\Delta_t^{\mathrm{GAS}}\right), \\ \Delta_t^{\mathrm{GAS}} &= \frac{1}{n} w_t \left[(y_t^\top \iota)^2 \left(\frac{1}{1+(n-1)\rho} + \frac{\rho}{1-\rho}\right) - (y_t^\top y_t)\frac{1+(n-1)\rho}{1-\rho}\right] + (n-1)\rho. \end{aligned} \tag{3.30}$$

If the GAS model is correctly specified, we can make use of Bartlett's identity such that $\mathbb{E}_{t-1}[\Delta_t^{\mathrm{GAS}}] = 0$ at the true parameters. The relevant condition for covariance stationarity is $|\beta| < 1$; see also Section 3.3.

The updating mechanism in (3.29) is markedly different from that in (3.30). First,

Δ_t^{GAS} exploits more information in the score via the inclusion of the weights w_t in (3.30). Second, Δ_t^{GAS} not only reacts to $(\iota^\top y_t)^2$, but also to the sum of squares $y_t^\top y_t = \|y_t\|^2$. As an example, consider the case of the uncorrelated normal distribution with $w_t = 1$ and $\rho = 0$. In this case, Δ_t^{GAS} reduces to

$$\Delta_t^{\text{GAS}} = \frac{1}{n^2}\iota^\top \left[n\, y_t y_t^\top - \frac{y_t^\top y_t}{n} \iota\iota^\top \right] \iota = \frac{(\iota^\top y_t)^2 - n}{n} - \left(\frac{y_t^\top y_t}{n} - 1 \right). \tag{3.31}$$

We see that the GAS model penalizes shocks with large $\|y_t\|$ by imposing a negative impact of $\|y_t\|$ on ρ_{t+1}. In fact if the elements of y_t are correlated and have the same standard deviation, one expects the individual elements y_{it} to be of similar magnitude, while pairwise deviations result in a correction of the correlation parameter through the updating scheme (3.30). Also the DCC-DECO model accounts for the squared observations y_{it}^2, but in a dynamic manner. A larger value for $\|y_t\|$ relates to (at least some of) the diagonal entries of Q_{t+1} to *increase*. By the transformation from Q_{t+1} to R_{t+1}^{DCC}, this typically results in a lower subsequent equicorrelation as well. For the DCC-DECO there is also the additional effect in the next period, as the larger diagonal elements of Q_{t+1} are carried forward via the recursion of Q_t. In the end, the dissipation of the effect of a large $\|y_t\|$ depends on the entire dynamic pattern of $\tilde{Q}_{t+1}^{-1/2}\tilde{Q}_t$, which is complicated to pin down given the dynamics of Q_t. Also note that in higher dimensions, the GAS transition dynamics in (3.30) put more emphasis on the cross-products $(\iota^\top y_t)^2 - n$ and less on the magnitude $\|y_t\|^2$.

Next to the above more subtle differences in the correlation dynamics of the GAS-DECO and DCC-DECO models, these models also have a number of more apparent differences. First, due to the univariate modeling of the time-varying correlation parameter in the GAS model we no longer require a correlation targeting procedure to facilitate the estimation of the intercept $\bar{Q}$ in (3.6). The estimation of $\bar{Q}$ in high cross-sectional dimensions is not only statistically inefficient, but introduces an incidental parameter problem that can render the maximum likelihood estimator inconsistent for the full DCC model; see Engle et al. (2008). In such cases, alternative estimation methods such as composite maximum likelihood estimation may be considered.

Second, unlinke the full recursion (3.28), the score-based correlation dynamics can be written in terms of univariate sufficient statistics. This avoids a high-dimensional $n \times n$ recursion in the auxiliary matrices Q_t. Pre-computing the sufficient statistics for the score-based recursion allows us to obtain a highly numerically efficient model and estimation procedure compared to the already efficient DCC-DECO methodology. In arbitrarily high dimensions, the estimations and recursions will not outpass the complexity of a simple univariate GARCH models. By contrast, the original GAS approach for full correlation matrices as advocated in Creal et al. (2011) requires the calculation of inverses and multiplications of $n^2 \times n^2$ matrices, which quickly becomes infeasible for large n.

Third, as the likelihood can be computed only via the univariate sufficient statistics, this implies that we can allow the dimension of observables to vary over time. If

the time-varying dimensionality is not driven by systematic sample selection issues, we can accomodate unbalanced panels and randomly missing observations. This is relevant for instance in our application to S&P 500 stock returns in Section 3.5.[4]

Finally, the reparameterization $h(f_t)$ in the score-driven approach allows us to easily include exogenous variables into the correlation dynamics, while retaining the equicorrelation structure and the positive definiteness of the correlation matrix. By contrast, the DCC-DECO dynamics do not ensure that the lower bound $-(n-1)^{-1}$ on the admissible range of equicorrelations is automatically satisfied, particularly not if also exogenous variables are included. Though in-sample problems for the DCC-DECO are expected to be small as the QML criterion penalizes violations of the lower admissible region, out-of-sample forecasts and risk analyses might be much more problematic. No such difficulties are expected for the GAS-DECO set-up of the model.

3.3 Statistical Properties of the GAS-DECO Model

Notation

We summarize the notation for some frequently used operators, which are also applicable

[4]See also Patton (2006a) for a discussion of copula estimation in overlapping panels.

for the rest of this thesis. We denote the natural filtration by $\mathscr{F}_t = \sigma(Z^t), t \in \mathbb{Z}$, where $Z^t = \{y_t, y_{t-1}, \ldots\}$ and y_t denotes the vector of observables. $A \otimes B$ denotes the Kronecker product of matrices A and B (see Abadir and Magnus (2005)); $\cdot$ denotes matrix multiplication and $\log(\cdot)$ denotes the natural logarithm. $\text{vec}(A)$ is the vec-operator stacking the columns of A and $\text{vech}(A)$ does the same for lower diagonal elements of A; $\mathcal{D}$ denotes the duplication matrix such that $\mathcal{D}\text{vech}(A) = \text{vec}(A)$; $A^\top$ denotes the transpose of a matrix ; A^+ denotes its Moore-Penrose inverse; $\|A\|$ denotes its generic norm, depending on the problem at hand we often make use of Frobenius norm $\|\cdot\|_F$, which is obtained by vectorizing the columns of A or the Euclidian norm $\|\cdot\|_2$, which is defined as the square root of the maximal eigenvalue of $A^\top A$. For a random variable A, we define the L^r-norm $\|A\|_r = \mathbb{E}\left[\|A\|^r\right]^{\frac{1}{r}}$, where $\mathbb{E}[\cdot]$ and $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \,|\, \mathscr{F}_t]$ denote the expectation and conditional expectation operators, respectively. $A_{i,j}$ denotes the (i,j)-th element of A, $A_{i,\cdot}$ denotes its i-th row and $A_{\cdot,j}$ denotes its j-th column. We denote the vector of ones $\iota_n \in \mathbb{R}^{n\times 1}$ interchangeably also as ι. Similarly, I_n denotes the unit matrix. Finally, the abbreviation "i.i.d." denotes independently, identically distributed observations.

3.3.1 Stationarity and ergodicity of the simulated model

Using Proposition 1, we can rewrite the scaled score as $s(f_t) = K(f_t)m_t$, with $K(f_t) = S(f_t)\Psi(f_t;\lambda)^\top$, and $m_t = \text{vec}\left(-\nabla_{p_{u,\lambda}}(u_t)u_t^\top - \mathrm{I}_n\right)$. Note that m_t is a mar-

tingale difference sequence, which is also required to asymptotic properties of the ML estimator. Let $\mathcal{F}_t$ and Θ denote sets where the time-varying parameter f and the static parameter θ take values, respectively. We obtain the following result for the known, correctly specified data generating process.

Proposition 2. *If the score of GAS model satisfies*

$$s(f_t) = K(f_t; \lambda)\, m_t, \tag{3.32}$$

with $\{m_t\}_{t\in\mathbb{Z}}$ *an i.i.d. sequence with zero mean and finite moment* $\mathbb{E}\,\|m_t\|^r < \infty$ *for some* $r > 0$*;* $K : \mathcal{F} \times \Lambda \to \mathbb{R}^{k\times n^2}$ *is continuous and satisfies*

$$\sup_{f\in\mathcal{F}} \|K(f;\lambda)\|_2 \leq \overline{K}(\lambda) < \infty, \tag{3.33}$$

for every $\lambda \in \Lambda$*. Furthermore, assume that there exists* $k \in \mathbb{N}$ *such that the* k*-step transition density is strictly positive and continuous, i.e.*

$$\mathbb{P}(f_{t+k} \in A | f_t = f) = \int_A \mu(y|f)dy \text{ and } \mu(\cdot|f) > 0 \quad \text{for all } t \in \mathbb{N}, \text{ all } f \in \mathcal{F} \text{ and all } A \in \mathcal{B}(\mathcal{F}), \tag{3.34}$$

where μ *denotes the Lebesgue measure.*

Then for all β *with modulus less than one and full-rank* α*, the chain* $\{f_t\}$ *is geometrically ergodic and the non-anticipative strictly stationary solution of* (3.8)–(3.9) *is geometrically* β*-mixing. If* $\mathbb{E}\|m_t m_t^\top\| < \infty$*,* $\{f_t\}$ *is also covariance stationary.*

Recall from Example 1 in Section 3.2 that in the bivariate Gaussian correlation model, we have that $K(f_t;\lambda) = (\sqrt{1-\rho(f_t;\lambda)^2}\,,\, \rho(f_t;\lambda))$ and $m_t = m(u_t) = (u_{1t}u_{2t}\,,\, 1-u_{2t}^2)^\top$. The vector $K(f_t)$ in this case lies on the unit circle and therefore always has full rank. Furthermore, condition (3.34) is satisfied in Gaussian models for single step transitions ($k = 1$) as the score driven innovations enable unlimited steps across the state space. In order to establish the condition (3.34) for scores that introduce bounded innovations, it is necessary to consider several chain transitions ($k > 1$) in order to ensure that the Markov chain does not get "stuck" without having access to all subsets of the entire state space $\mathcal{F}$. For such models, the irreducibility property needs to be established on a case-by-case basis, either due to further analytical simplifications or numerical procedures such as the stability regions discussed in Chapter 2.

The statements in Proposition 2 not only apply to processes from the equicorrelation model, but to the wider class of multivariate scale models with a bounded scaling function $K(f_t;\lambda)$. As such, the results of this proposition can be used for correlation models with multiple time varying correlations. Also note that the geometric ergodicity property entails that the simulated paths converge to the stationary distribution at an exponential exponential rate.[5]

[5]A Markov chain $\{f_t\}_{t=0,1,2,\ldots}$ is said to be geometrically ergodic if there exists some constant $0 < s < 1$ such that the Markov chain n-step transition probability $\mathbb{P}^n(\cdot|f)$ and the stationary distribution $\pi(\cdot)$ satisfy the relationship

$$\|\mathbb{P}^n(\cdot|f) - \pi(\cdot)\| \leq Ms^n \qquad \text{for all } f \in \mathcal{F}.$$

The boundedness of $K(f_t;\lambda)$ depends on the boundedness of the product of the scaling function $S(f_t;\lambda)$ and the Jacobian $\Psi(f_t)$ in Proposition 2. For GAS models that use particular forms of information matrix scaling as in equation (2.5), we can establish boundedness of $K(f_t;\lambda)$ and thus stationarity and ergodicity (SE) as a direct consequence of the above result. This is stated in the following two corollaries.

Corollary 2. *Consider the model described by equations* (3.8)–(3.11) *with scaling matrix* $S(f_t;\lambda) = (\mathcal{I}(f_t;\lambda))^{-a}$ *and* $\text{rank}(\Psi(f_t)) = k \leq \text{rank}(\mathcal{I}_{p_{u,\lambda}} - \text{vec}(\text{I}_n)\text{vec}(\text{I}_n))$. *If either of the following two conditions holds,*

(i) $a = 1/2$*;*

(ii) $a \in \{0, 1\}$*, and* $|h(f_t;\lambda)^{-1}_{i,j}\partial h_{r,s}(f_t;\lambda)/\partial f_l| \leq K < \infty$ *for all* $i, j, r, s = 1, \ldots, n$, $l = 1, \ldots, k$*, and all* $f \in \mathcal{F}$*;*

then (3.33) *is satisfied. If also the other conditions of Proposition 2 hold, there exists a stationary and ergodic solution* f_t *to equations* (3.8)–(3.11)*.*

Corollary 3. *The Gaussian GAS equicorrelation model of Corollary 1 with unit scaling function* $S(f_t;\lambda) \equiv 1$ *has a stationary and ergodic solution if* β *has modulus less than one. Furthermore, for a Student's t model, if the chain* $\{f_t\}$ *is psi-irreducible, then stationarity can be ensured under the same assumptions.*

The simple structure of the GAS-DECO model enables us to establish a straightforward intuition for the above results. For instance, in the case of Example 1 in Section 3.2 we saw for the bivariate normal distribution that we can write the score

as $K(f_t;\lambda) = (\sqrt{1-\rho(f_t;\lambda)^2}\,,\,\rho(f_t;\lambda))$ and $m_t = m(u_t) = (u_{1t}u_{2t}\,,\,1-u_{2t}^2)^\top$. The vector $K(f_t)$ in this case lies on the unit circle and therefore always has full rank. Corollary 3 then ensures a stationary and ergodic solution f_t exists if $\|\beta\| < 1$. For the adjusted tanh parameterization in equation (3.17) and unit scaling $S(f_t;\lambda)$, we also obtain that $K(f_t;\lambda)$ is bounded, though we can no longer ensure that $K(f_t;\lambda)$ always has full rank.

The nature of the current conditions differs from conditions in Chapter 2. From a stochastic process perspective, the stationarity conditions are less strict. A particularly appealing feature of the current approach is its invariance to the dimensions of y_t and f_t, the fat-tailedness of u_t, and observationally equivalent decompositions of y_t into u_t. On the other hand, by having imposed the stricter conditions in the previous chapter, we had additionally gained uniqueness and convergence properties of the stochastic recurrence equations.

A theoretical investigation of the necessary conditions is beyond the scope of this paper as standard tools for characterizing explosiveness as $\|\beta\| \geq 1$ due to Nelson (1990) and Klüppelberg et al. (2004) make use of the linear updating scheme of the GARCH model; see also Kleibergen and van Dijk (1993) and Francq and Zakoïan (2012). This contrasts sharply with the nonlinear structure of typical GAS recursions.

Finally, we establish moments of the stationary solution in the following proposition.

Proposition 3. *Under the assumptions of Corollary 2, if* $\mathbb{E}\|m_t\|^r < \infty$*, then* $\mathbb{E}\|f_t\|^r < \infty$*.*

Proposition 3 establishes the existence of moments of f_t as generated by the non-linear GAS recursion using the uniform boundedness of $K(f_t; \lambda)$ and the existence of moments of the i.i.d. sequence m_t.

3.4 Simulation evidence

We compare the ability of the GAS-DECO and DCC-DECO model to extract a time-varying correlation parameter in cases where neither model is correctly specified. As in Engle (2002a), we consider deterministic correlation patterns in the data generating process and extract the time-varying correlations via the GAS-DECO and DCC-DECO filters. We report the results for four different patterns: a single large break, a trigonometric sine function, an exogenous AR(1) process, and a local level (LL) model. We consider the Student's t distribution with λ degrees of freedom for the error density p_u. The results are summarized in Table 3.1 and Figure 3.4 for cross-sectional dimensions $n \in \{2, 10, 100\}$ and fat-tailedness of the observation density $\nu \in \{5, \infty\}$.

Table 3.1 reveals that the score-driven equicorrelation approach outperforms the

Table 3.1: Signal Extraction Results

Notes: We present the mean squared error (MSE), median absolute deviation (MAD), and likelihood value based on a multivariate Gaussian observation density (QLIK) for different data generating processes. Each process is simulated 100 times. The GAS-DECO model is estimated with unit scaling ($a = 0$). The different data generating processes are visualized in Figure 3.4.

(a) Gaussian noise

		$n = 2$			$n = 10$			$n = 100$		
		MSE	MAD	QLIK	MSE	MAD	QLIK	MSE	MAD	QLIK
Break	GAS	0.0675	0.0387	−2551	0.0409	0.0198	−10300	0.0235	0.0102	−90727
	DCC	0.0745	0.0430	−2553	0.0564	0.0295	−10327	0.0529	0.0267	−91179
Sine	GAS	0.1189	0.0742	−2554	0.0592	0.0357	−10359	0.0330	0.0204	−91554
	DCC	0.1299	0.0798	−2560	0.0831	0.0529	−10441	0.0757	0.0463	−92693
AR(1)	GAS	0.1580	0.1110	−2575	0.1211	0.0806	−10434	0.0984	0.0656	−91584
	DCC	0.1606	0.1133	−2578	0.1414	0.0997	−10552	0.1366	0.0965	−93658
LL	GAS	0.1040	0.0699	−2729	0.0621	0.0394	−12467	0.0403	0.0252	−116301
	DCC	0.1089	0.0726	−2730	0.0761	0.0480	−12485	0.0696	0.0438	−116613

(b) Student's $t(5)$ noise

		$n = 2$			$n = 10$			$n = 100$		
		MSE	MAD	QLIK	MSE	MAD	QLIK	MSE	MAD	QLIK
Break	GAS	0.0749	0.0482	−2563	0.0479	0.0258	−10330	0.0403	0.0207	−90788
	DCC	0.0953	0.0578	−2566	0.0692	0.0393	−10362	0.0643	0.0353	−91249
Sine	GAS	0.1287	0.0774	−2560	0.0744	0.0441	−10382	0.0615	0.0374	−92454
	DCC	0.1558	0.0990	−2573	0.1003	0.0610	−10479	0.0900	0.0565	−93453
AR(1)	GAS	0.1622	0.1108	−2578	0.1321	0.0865	−10477	0.1275	0.0814	−93140
	DCC	0.1733	0.1237	−2581	0.1520	0.1081	−10573	0.1475	0.1047	−94051
LL	GAS	0.1088	0.0729	−2741	0.0698	0.0451	−12475	0.0610	0.0388	−116188
	DCC	0.1377	0.1004	−2742	0.0897	0.0566	−12491	0.0839	0.0522	−116416

DCC-based updating scheme in terms of root mean squared error (MSE), median absolute deviation (MAD) and likelihood-based (Kullback-Leibler divergence) loss functions (QLIK) for all dimensions n, data generating processes, and distribution choices.

Figure 3.4 moreover shows that the relative performance of the GAS-DECO filter

depends on the informativeness of the likelihood. For instance, if the data is more fat-tailed ($\lambda = 5$), the signal in the score is weaker and signal extraction results are noisier. This can be seen from the wider grey areas in the figure for the Student's $t(5)$ compared to the Gaussian distribution. Similar effects (not shown) can be observed when changing the cross-sectional dimension and the level of correlation. These properties are conveniently summarized by the properties of the information matrix. It is straightforward to show that in the case of the GAS-DECO $\partial \mathcal{I}_t / \partial n > 0$, $\partial \mathcal{I}_t / \partial \nu > 0$ and $\partial \mathcal{I}_t / \partial |\rho| > 0$. A larger expected curvature of the likelihood as described by the conditional Fisher information matrix implies a more precise signal extraction.

3.5 Empirical applications

3.5.1 S&P500 equity returns

In this section we study the time-varying equicorrelation between the constituents of the S&P 500. We use data stock return data from January 1, 2000 to December 31, 2012 for the companies listed in Table 3.6 in the appendix of this chapter. To fully focus on the correlation part of the model, we first filter all series using a GJR(1,1)-model.[6] In order to keep the GAS-DECO and DCC-DECO comparable, we use the same marginal filter for both models rather than using a GAS based filter for the GAS-DECO, and a GJR filter for the DCC-DECO. Having obtained the standard-

[6]The results are robust to using an EGARCH(1,2) model for the marginal volatilities.

ized residuals from this marginal filtering step, we model the subsequent dependence dynamics using the DECO specification.

Note that this two-step procedure is possible due to the use of the Gaussian QML criterion for estimation. Other criterion functions typically cannot be decomposed into two separate terms relating to volatilities and correlations, respectively. If the criterion cannot be decomposed in this way, the computational appeal of the two-step procedure is lost.

The role of the Student's t degrees of freedom parameter ν is subtle and requires further discussion. In the full dataset we estimate the average marginal degrees of freedom in the filtered residuals to be approximately 6. Contrasting to the focus of Creal et al. (2011) on density forecasting, this paper has followed a QML approach to estimate the latent dynamic equicorrelation. Thus the observation density and the score-based updating rule are not necessarily linked in the current approach. In fact, as the equicorrelation assumption misspecifies the model to favor parsimony, it is desirable to make the innovations to the time-varying parameter robust to the implications of the misspecified likelihood. The signal to noise in GAS models is jointly determined by ν and α in the score based updating scheme. To avoid possible identification problems in finite samples, we fix $\nu = 5$ to mimic a low signal-to-noise ratio in economic data. This has worked well across a wide variety of applications and data sets, yet with such a high degree of tail fatness we typically find in equity returns, skewness in the distribution has a less pronounced effect for density forecasting

(Creal and Tsay, 2014; Lucas et al., 2014).

Figure 3.1 displays the filtered correlations based on QML estimation. Table 3.2 summarizes these results. Compared to the DCC formulation of Engle and Kelly (2012), we see in fact quite similar patterns. The equicorrelation is relatively moderate in the beginning of the 2000's and increases clearly in times of time crisis. The pronounced difference is in periods where the correlation goes up for short episodes. The DCC formulation is more volatile and gives almost always somewhat higher equicorrelation point estimates. This is in agreement with the discussion Example 1, demonstrating that the score includes a penalty term induced by tail dependence.

As a diagnostic device to explore model uncertainty among different equicorrelation models, we also compare the GAS filtered paths with the particle filter output, where the score-driven innovations in the state equation are replaced with orthogonal NID $(0, 1)$ shocks and static hyperparameters are obtained from ML estimation of the observation-driven GAS model.[7] It striking that both filtered paths largely coincide, supporting the closeness of parameter driven models and GAS observation-driven models simulations on real economic data.[8] As a benefit, the confidence bands around the paths of the parameter driven model reveal that the equicorrelation truly is time-varying, and the differences between the DCC-DECO and GAS models substantially differ in single episodes where the DCC-DECO-filtered equicorrelation shoots up.

[7] In order to ensure homoskedasticity of the innovations in both models, we implement the equicor-

Table 3.2: QML estimation of equicorrelation models

Notes: The GAS model is estimated with unit information scaling ($a = 0$) due to the variance stabilizing Fisher transformation. The DCC column follows the approach by Engle and Kelly (2012). The analysis is conducted for the period between January 1, 2000 and December 31, 2012. The included 154 stocks are reported in Table 3.6.

	DCC-Deco QMLE	t(5)-GAS-Deco QMLE	$t(\nu)$-GAS-Deco QMLE	$t(\nu)$-GAS-Deco MLE
ν DoF		5.0000	38.2290	4.9717
			(0.0324)	(0.0020)
c		0.0073	−0.0612	−0.0094
		(0.0000)	(0.0000)	(0.0000)
α	0.0528	0.0303	0.0263	0.0078
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
β	0.9411	0.9950	0.9224	0.9818
	(0.0001)	(0.0002)	(0.0009)	(0.0002)
LogL	−636298	−635283	−633437	−614744
AIC	1272602	1270573	1266883	1229496
BIC	1272621	1270591	1266908	1229520

Both models have quite similar economic implications in terms of the extracted equicorrelation. It is interesting to observe that the relative performance of both models varies systematically over time as documented by the cumulative quasi-likelihoods plot in Figure 3.1. In tranquil times, both models have a comparable performance, whereas periods with the presence of outliers favor the robust ($\nu = 5$) GAS specification. This feature of the data can be incorporated in a more refined regime-switching multivariate models of Haas et al. (2004) and Boudt et al. (2012). We further note that the GAS equicorrelation model can be in principle estimated without imposing the stationarity constraints as they are typically satisfied when implementing an

relation model with square root information matrix scaling ($a = 1/2$).

[8] See also Koopman et al. (2012) for a simulation study under a controlled setting.

unconstrained optimization procedure.

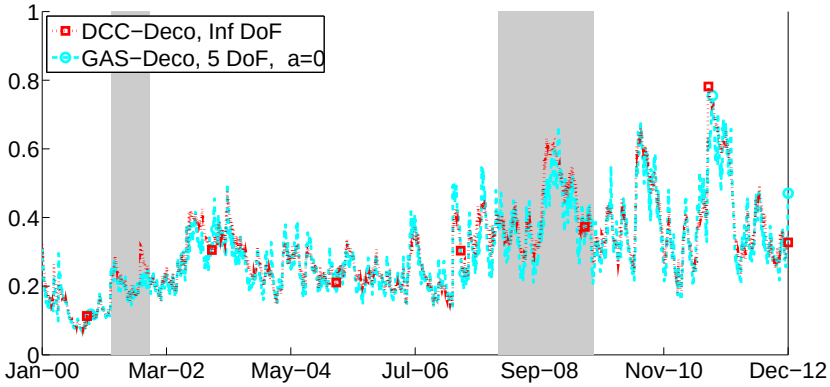
3.5.2 A joint model to extract information from options data

A central question in financial economics addresses the information content in traded options. As option prices are determined by investor expectations (and possibly also by preferences), one would expect that options data would provide the econometrician with valuable information regarding the latent dynamics of the underlying asset.

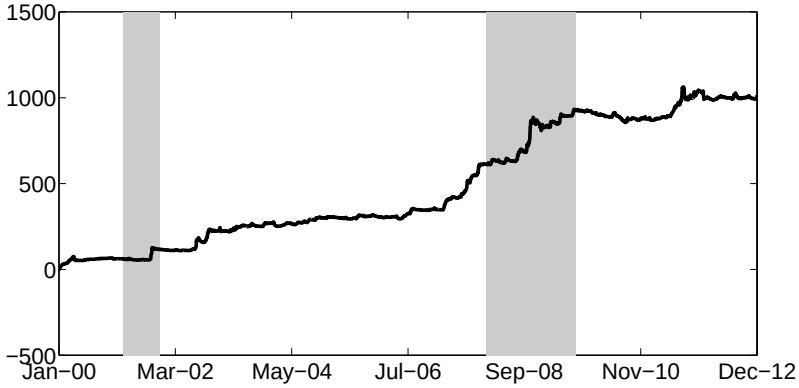
Recent literature on risk sharing and portfolio choice considers the impact of uncertain diversification opportunities on aggregate and individual decisions, see also **?** and Christoffersen et al. (2012). Driessen et al. (2009) develop a partial equilibrium model where the fear of deterioration of diversification opportunities is compensated in equilibrium. More specifically, they interpret the difference between risk-neutral and physical correlations as the correlation risk premium. Similarly, Buraschi et al. (2010) develop a general equilibrium model with endogenous time-varying correlation risk premia. They find that both theoretically and empirically, correlation risk premia increase steeply in economic downturns, moreover it is correlation risk premia and not necessarily variance risk premia, which are priced in equilibrium. This result is complemented by Merton's ICAPM model, in which the hedging demand against

correlation risk is found to be of greater magnitude than hedging motives for volatility risk.

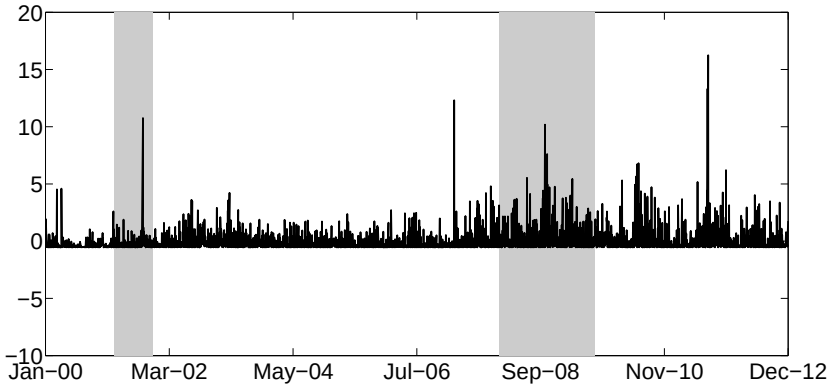
In this application, we are less interested in the full multivariate density of individual stock index constituents and focus on the underlying correlation levels and risk premia estimates. In order to avoid the adverse effects of mis-specification to such a high-dimensional observation density, we apply the QML estimation routine, which ensures consistent estimates of the static model parameters even when the observation density is unknown. In doing so, we depart from the finance literature on correlation risk premia by specifying a full joint model for options and returns data.



(a) Filtered correlations



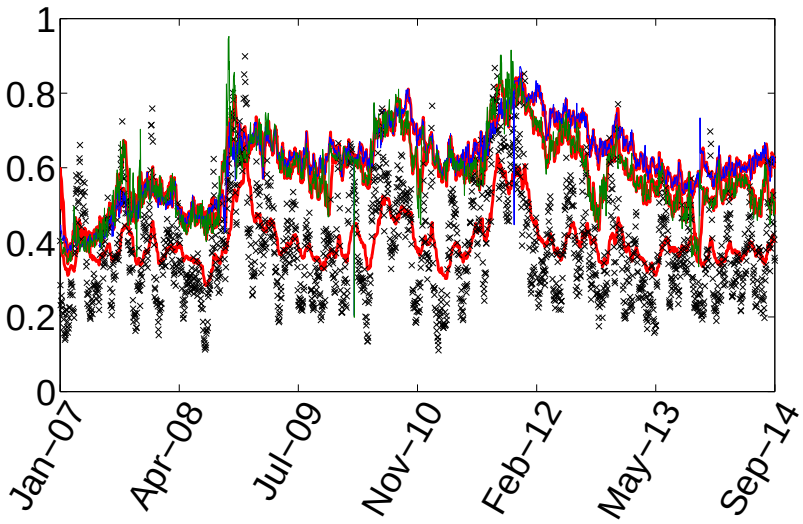
(b) Cumulative differences in likelihoods (GAS-DCC)



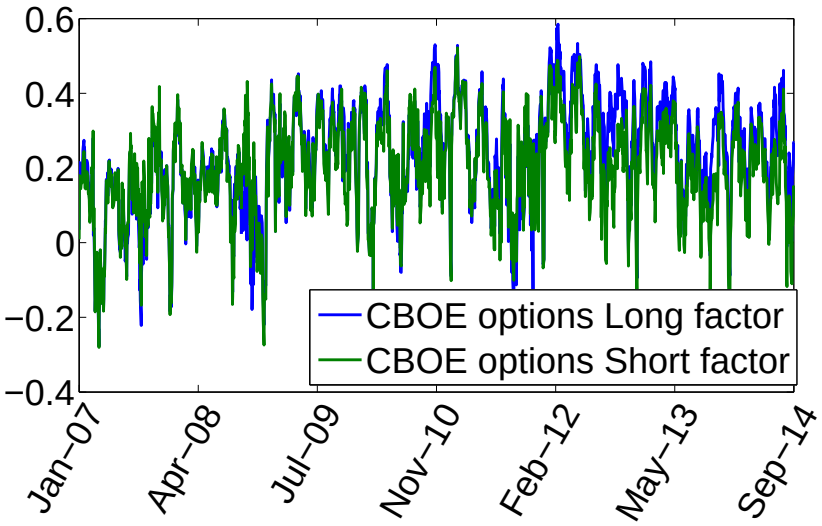
(c) Evolution of $\{y_t^\top y_t\}$

Figure 3.1: Comparison of equicorrelation models

Notes: Based on QML estimation with Student's t score and $n = 154$. The estimation results are summarized in Table 3.2. The shadowed areas denote periods of NBER recessions.



(a) Fit of the model



(b) Correlation risk premia

Figure 3.2: Correlation path estimates

Notes: In panel (a), the red lines depict filtered time-varying parameters. The blue (green) lines reflect long(short)-term options data. The black crosses depict GAS-extracted physical correlations.

The data stems from the Chicago Board Options Exchange (CBOE), which, as of January 3, 2007, reports market-implied estimates of the average correlation of representative stocks comprising the S&P 500 Index (SPX). Using SPX options prices, together with the prices of options on the 50 largest stocks in the S&P 500 Index, the CBOE S&P 500 Implied Correlation Indexes offers insight into the relative cost of SPX options compared to the price of options on individual stocks that comprise the S&P 500. CBOE calculates and disseminates two indexes tied to two different maturities, usually one year and two years out, with rotating tickers: ICJ, JCJ and KCJ. The underlying options have a *fixed* maturity (January of the following year and the year thereafter). An overview of the included stocks is reported in Table 3.6 in the Appendix.[9]

We can calculate the option-implied variance of a portfolio as the weighted sum individual option-implied volatilities

$$\left(\sigma_{SPX}^{IV}\right)^2 = \sum_{i=1}^{N} w_i^2 \left(\sigma_i^{IV}\right)^2 + \sum_{i \neq j} w_i w_j \rho_{ij} \sigma_i^{IV} \sigma_j^{IV} \tag{3.35}$$

and by restricting the pairwise correlations to be identical (i.e., $\rho_{ij} \equiv \bar{\rho}$) we have

$$\bar{\rho} = \frac{\left(\sigma_{SPX}^{IV}\right)^2 - \sum_{i=1}^{N} w_i^2 \left(\sigma_i^{IV}\right)^2}{\sum_{i \neq j} w_i w_j \rho_{ij} \sigma_i^{IV} \sigma_j^{IV}},$$

[9] A practical issue one needs to deal with is that in raw data, the implied correlations are not always bounded by one in absolute value. In the full sample we have 3 degeneracy occasions in short term implied correlations during the peak of the financial crisis, more specifically on 06-Nov-2008, 13-Nov-2008, and 20-Nov-2008, we treat these observations as missing, as is common in the state space framework.

where σ_{SPX}^{IV} and σ_i^{IV} reflect index implied and individual stock implied volatilities, respectively. These are obtained by inverting the Black-Scholes option pricing formula for at-the-money options. We illustrate the individual weights of the $N = 50$ stocks in Table 3.6 in the appendix of this chapter for a given date of May 29, 2009. These weights vary of course over time, and are accounted for by the time series of $\bar{\rho}$ as obtained directly from CBOE.

Furthermore, the average correlation can also be traded in practice. As the index volatility is increasing in $\bar{\rho}$ for given values of w_i and σ_i^{IV}, one can bet on future implied correlations by means of a long-short trading strategy that involves taking opposite positions in index and individual asset volatility derivatives such as straddles or variance swaps.

The model developed in this paper enables us to further study time-varying correlation risk premia, which are defined as the difference between risk neutral and physical correlation parameters. We adopt a measurement error perspective in a dynamic joint model for returns and option-implied data. We define a structural time series model by augmenting the $\tanh_n^*$-parametrized equicorrelation system (3.8)–(3.11) with the measurement equations for short term and long term option-implied correlations (with maturities in the following year for ρ_t^{Short} and in the year thereafter for ρ_t^{Long}, respectively). The additional measurement equations and recurrence

equations for time-varying parameters are formulated by

$$\tanh_n^{*-1}\left(\bar{\rho}_t^{Short}\right) = \gamma^{Short} + f_t^{Short} + f_t^{Corr} + \varepsilon_t^{Short}, \quad \varepsilon_t^{Short} \overset{i.i.d.}{\sim} (0, \sigma^2_{Short}) \tag{3.36}$$

$$\tanh_n^{*-1}\left(\bar{\rho}_t^{Long}\right) = \gamma^{Long} + f_t^{Long} + f_t^{Corr} + \varepsilon_t^{Long}, \quad \varepsilon_t^{Long} \overset{i.i.d.}{\sim} (0, \sigma^2_{Long}) \tag{3.37}$$

and specifying the factor dynamics by

$$\begin{pmatrix} f_{t+1}^{Short} \\ f_{t+1}^{Long} \\ f_{t+1}^{Corr} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \omega \end{pmatrix} + \begin{pmatrix} \beta_1 & 0 & 0 \\ 0 & \beta_2 & 0 \\ 0 & 0 & \beta_3 \end{pmatrix} \begin{pmatrix} f_t^{Short} \\ f_t^{Long} \\ f_t^{Corr} \end{pmatrix} + \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix} \begin{pmatrix} \nabla_t^{Short} \\ \nabla_t^{Long} \\ \nabla_t^{Corr} \end{pmatrix}, \tag{3.38}$$

where ∇_t^{Long} and ∇_t^{Short} stem from location models (3.36) and (3.37) for inverse $\tanh_n^*$-transformed option-implied correlations and are calculated from a Student's t density with $\nu^{Options}$ degrees of freedom. For assets' equity returns, the role of the usual equicorrelation update ∇_t^{Corr} was discussed earlier in this paper. The physical correlations are modeled by a mean-reverting process as discussed in the previous section.[10] The long-run correlation risk premia are captured by γ^{Short} for short-dated options and by γ^{Long} for long-dated options. Furthermore, changes in the short-dated and long-dated correlation risk premia are captured by the respective time-varying

[10]One could extend this model by specifying parametric models to describe expectation of physical correlations at option maturity (see also van Dijk et al. (2014)). However, we do not find an improvement in considering the standard functional forms that are derived from the notion of a mean-reverting physical correlation process.

Table 3.3: Parameter estimates for the dynamic correlation risk premia model

Parameter	α_1	α_2	α_3	β_1	β_2	β_3	ω	γ^{Short}	γ^{Long}	$1/\nu^{Options}$
Estimate	0.0704	0.0559	0.0069	1.0044	1.0056	0.9957	0.0012	0.3486	0.4129	0.4818

parameters f_t^{Short} and f_t^{Long}. Such a parsimonious specification serves to capture the basic features of option-implied correlation data, namely correlation risk premia and time-variation thereof.

The model is estimated by maximizing the joint quasi-likelihood function corresponding to the measurement equations in (3.8), (3.36) and (3.37). The parameter estimates for the joint model are given in Table 3.3, while the filtered time-varying parameter paths are presented in Figure 3.3. Furthermore, Figure 3.3 extracts the dynamic risk premia. The model-implied risk-premium, i.e., the average absolute difference between correlations under physical and risk-neutral measure (captured in a non-linear manner by the parameter γ^{Short}), is around 0.20, which is in line with prevailing estimates in the literature (cf. Buraschi et al., 2010).[11] There are several outliers in the options data, such that the estimated Student's t degrees of freedom parameter is close to two.[12] As the model is fitted jointly on both equity returns and option price data, the dynamic physical correlation paths are also partly deter-

[11]The unconditional risk premia are inferred as the average difference between the fitted physical correlations and option-implied correlations. Alternatively, a model-based quantity to describe correlation risk premium can be computed as $\tanh_n^*\big(\omega/(1-\beta_3)+\gamma\big) - \tanh_n^*\big(\omega/(1-\beta_3)\big)$.

[12]The dates with notable spikes in long term option-implied correlations are: a) 03-Mar-2010, b) 21-Nov-2011, c) 20-Nov-2013, which are in the vicinity of the following macroeconomic events: a) The U.S. Treasury reiterates its commitment to protecting holders of debt in Fannie Mae and Freddie Mac (05-Mar-2010); b) Fed issues final rule on annual capital plans & Fed launches 2012 review, i.e. stress test (22-Nov-2011); c) JPMorgan agrees $13bn settlement with U.S. over mortgages (19-Nov-2013).

mined by the fit on options data. Smoothness of the jointly extracted paths suggests that the noisiness of the time-varying parameter paths is reduced by incorporating additional measurement equations. Large changes in dynamic physical correlations will also influence the fit of the option implied correlation measurements in (3.36) and (3.37), which favors smaller absolute values of physical correlation innovation parameter α_3 such that the extracted physical correlation paths will be smoother. Furthermore, in an unreported analysis via Nyblom tests, we do not find remaining time-variation in short term option residuals, but the long term options seem more volatile, possibly due to liquidity issues in long-maturity options, which could make the market prices for individual options less reliable and therefore distort our market implied correlation data. Our measurement error perspective captures such liquidity issues by introducing an error term in the option-implied correlation measurements (3.36) and (3.37). This way, we can estimate the signal content of option market implied correlation data.

Another striking feature of the fitted model is that the shocks to option-implied correlation risk premia are *estimated* to have an explosive impact. On the other hand, we obtain an estimated near, but smaller than a unit root for the innovation for the physical dynamic correlation parameter ($\widehat{\beta}_3 = 0.996$). We see from Figure 3.3 that the option implied correlations exhibit much more persistence than the estimated path of physical correlations might also explain why the literature finds no predictive power of option-implied correlations for realized correlations (see also Buss and Vilkov, 2012;

Christoffersen et al., 2011; DeMiguel et al., 2013): the two series have different dynamic properties, implied correlations being the most persistent. The estimated explosive behavior of the implied correlation dynamics makes statistical inference hard in this specific setting as much of the asymptotic theory used to derive the standard errors makes use of the stationarity assumption. This is why I refrain from reporting standard errors in this case.

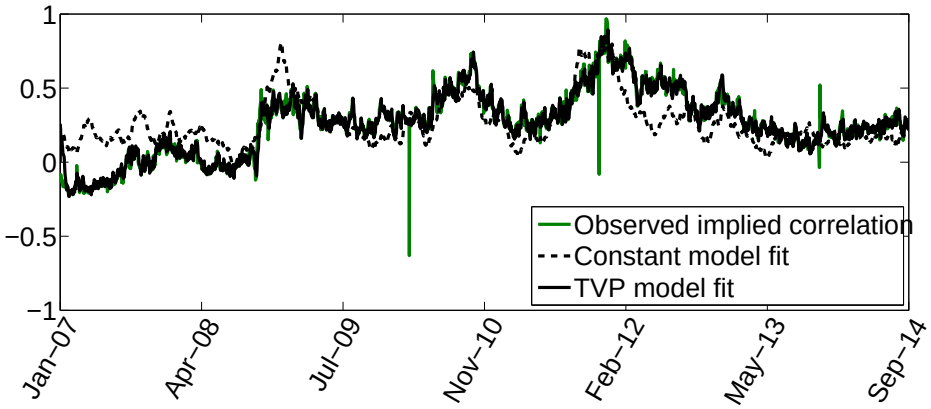
The time-varying nature of correlation risk premia is preferred over constant risk premia in terms of tests for remaining time variation in the model residuals. Table 3.4 conducts Nyblom tests on the inverse $\tanh_n^*$-transformed implied option correlations. We do indeed observe that the remaining time variation in residuals as measured by the Nyblom test statistic is reduced by introducing a time-varying parameter model. Furthermore, Figure 3.3 compares the option implied correlations to the model implied fit. We observe that constant risk premia impose parallel shifts of option risk premia, while the observed data can exhibit persistent deviations from this restriction.

Table 3.4: Nyblom tests for remaining time variation in residuals

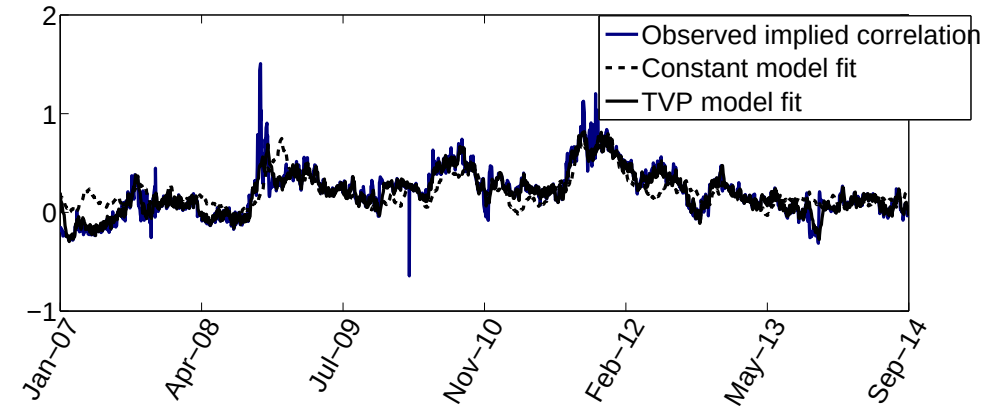
	Short term implied correlation		Long term implied correlation	
	Constant model	TVP model	Constant model	TVP model
Test Statistic	3.1405	0.4379	11.3297	1.2614
Reject at 5% level	Yes	No	Yes	Yes
Reject at 10% level	Yes	Yes	Yes	Yes

Notes: The test statistic refers to a Nyblom test with the null hypothesis of no remaining time variation in residuals.

The time-varying parameter model enables us to uncover dynamic features of option pricing markets. It is interesting to note that during the outbreak of the global financial crisis in 2008 and during the European debt crisis in 2011, the option market implied short term correlations were in an elevated state when compared to option implied long term correlations. A similar pattern has been established for implied volatilities, such that short-term implied volatilities tend to exceed long-term implied volatilities during periods of financial distress (see e.g. Koijen et al., 2015). On the other hand, according to the best knowledge of the author, dynamics of the correlation term structure have remained unexplored by the financial option pricing literature. In view of (3.35), there are two channels why index options become more expensive during crisis events. First, the volatilities of the underlying stocks are in relatively elevated levels with respect to their time series averages. Second, the short term market-expectation of correlations among the index constituents is relatively higher than the long term correlation expectations. In bull markets on the other hand, such as the period from 2012 until 2014, the long term option implied correlations exceed the short term counterparts, constituing an insurance risk premium or indeed a "correlation risk premium" for writing index options.



(a) Short term implied correlations



(b) Long term implied correlations

Figure 3.3: Comparison of constant versus time-varying parameter models

Notes: We take the inverse $\tanh_n^*$-transform of implied correlation as defined in the measurement equation (3.36) and (3.37), and compare the fit of the GAS time-varying correlation risk premia model (solid black line) with a constant risk premium model (dashed black line).

3.6 Conclusion

We recognize that score-driven correlation models give parametric weights to existing dynamic time-varying parameter updating schemes of Engle (2002a) and Patton (2006b). However, in higher dimensions the pre-existing dynamic innovations might become numerically cumbersome. Therefore, we introduced an alternative univariate updating mechanism, which enables a quick estimation routine in high dimensions, well beyond the cross-sectional dimension of 100 assets. Furthermore, the approach accommodates the estimation of dependence in unbalanced panels that may include missing observations. Empirically we corroborate the success of the dynamic equicorrelation model of Engle and Kelly (2012). Moreover, it is striking that the score-driven equicorrelation model gives nearly identical patterns when compared to a corresponding stochastic equicorrelation model obtained via particle filtering (see also Figure 3.5 in the appendix of this chapter). In a simulation study, the score-based approach outperforms the DCC approach in capturing the latent correlation patterns.

While having focused on modeling the dynamic equicorrelation, we have come across different theoretical properties of Generalized Autoregressive Score models of Creal et al. (2013). In short, in spite of the inherent non-linearities in correlation modeling, we still find that stationarity and ergodicity can be ensured under the relatively simple constraint that the autoregressive parameter is less than unity in absolute value.

There still remain open questions which we have not addressed by this paper.

Theoretically, one may delve deeper into the sharpness of the ergodicity conditions to establish simpler irreducibility properties for score driven models with fat-tailed innovations. We leave this for future work.

3.A Proofs

Proof of Proposition 1. Given the log density of the scale model

$$\log p_y(y_t|f_t;\lambda) = \log p_u(h(f_t)^{-1}y_t) - \log\det(h(f_t)),$$

we obtain

$$\begin{aligned}
\nabla &= \frac{\partial \log p_u(h(f_t;\lambda)^{-1}y_t)}{\partial f_t} - \frac{\partial \log\det(h(f_t;\lambda))}{\partial f_t} \\
&= \left(\frac{\partial \mathrm{vec}(h(f_t;\lambda))}{\partial f_t}\right)^\top \left(\frac{\partial (h(f_t;\lambda)^{-1}y_t)}{\partial \mathrm{vec}(h(f_t;\lambda))}^\top \frac{\partial \log p_u(u_t)}{\partial u_t} - \frac{\partial \log\det(h(f_t;\lambda))}{\partial \mathrm{vec}(h(f_t;\lambda))}\right) \\
&= \left(\frac{\partial \mathrm{vec}(h(f_t;\lambda))}{\partial f_t}\right)^\top \left(-\left(h(f_t;\lambda)\otimes h(f_t;\lambda)^\top\right)^{-1}(y_t\otimes \mathrm{I}_n)\nabla_{p_{u,\lambda}}(u_t) - \mathrm{vec}\left((h(f_t;\lambda)^\top)^{-1}\right)\right) \\
&= \left(\frac{\partial \mathrm{vec}(h(f_t;\lambda))}{\partial f_t}\right)^\top \left(\mathrm{I}_n\otimes h(f_t;\lambda)^{-1}\right)^\top \mathrm{vec}\left(-\nabla_{p_{u,\lambda}}(u_t)u_t^\top - \mathrm{I}_n\right) \\
&= \Psi(f_t;\lambda)^\top \mathrm{vec}\left(-\nabla_{p_{u,\lambda}}(u_t)u_t^\top - \mathrm{I}_n\right).
\end{aligned}$$

□

Proof of Proposition 2. In order to establish geometric ergodicity, we use the drift

conditions in Cline and Pu (1999) for recursions of the form

$$f_{t+1} = \alpha(f_t) + \gamma(u_t; f_t), \tag{3.39}$$

where u_t is an i.i.d. sequence of random variables, independent of the initial state f_0. This recursion encompasses the GAS(1,1) model by defining

$$\alpha(f_t) := \omega + \beta f_t \tag{3.40}$$

$$\gamma(u_t; f_t) := K(f_t)m(u_t). \tag{3.41}$$

We next restate the key conditions and verify their validity.

Assumption 1 (Assumption 2.1 of Cline and Pu (1999))**.** *Assume* $\{f_t\}$ *is an aperiodic,* ψ*-irreducible T-chain satisfying* (3.39)*. Let* $r > 0$ *and define* $\theta(f) = \frac{\alpha(f)}{1+\|f\|}$*. Assume that*

(i) α *is unbounded on* $\mathbb{R}^k$ *and* θ *is bounded on* $\mathbb{R}^k$*;*

(ii) $\sup_{\|f\|\leq M} \mathbb{E}\left(\|\gamma(u_1; f)\|^r\right) < \infty$ *for all* $M < \infty$*;*

(iii) $\lim_{\|f\|\to\infty} \mathbb{E}\left[\frac{\|\gamma(u_1;f)\|^r}{\|f\|^r}\right] = 0.$

Before verifying the stability conditions (i) – (iii), we demonstrate the requirements for Markov chain to satisfy the conditions for the setting of Cline and Pu (1999) to apply.

Lemma 5 (Lemma 1 of Bec et al. (2008))**.** *Under Assumption* (3.34) *the Markov chain* $\{f_t\}_t$ *is psi-irreducible and aperiodic.*

Due to the assumption (3.34), the required aperiodicity and psi-irreducibility follow directly from Lemma 1 in Bec et al. (2008). In order to confirm the definition T-Chain property, we need to verify that there exists a kernel T and a (possibly trivial) probability distribution $\{a_n\}$ on the nonnegative integers such that

(a) $T(f, \mathcal{F}) > 0$ for all $f \in \mathcal{F}$;

(b) $T(\cdot, A)$ is continuous for all $A \in \mathcal{B}(\mathcal{F})$;

(c) $\sum_{n=0}^{\infty} a_n P_n(f, A) \geq T(f, A)$ for all $f \in \mathcal{F}$, $A \in \mathcal{B}(\mathcal{F})$.

The properties (a)–(c) also follow from (3.34) as we can simply set $T(f, A) := \mathbb{P}(f_{t+k} \in A | f_t = f)$ and $\{a_n\} := \{\mathbf{1}[n = k]\}$.

We now validate the rest of assumptions

(i) See (3.40) and observe $\|\theta(f)\| \leq \|\omega\| + \|\beta\|$.

(ii) For all $M < \infty$ $\sup_{\|f\| \leq M} \mathbb{E}\|\gamma(m_1; f)\|^r \leq \overline{K}\ \mathbb{E}\|m_1\|^r < \infty$, where the moment condition on m_1 was assumed in the lemma.

(iii) By the majorant criterion,

$$\lim_{\|f\| \to \infty} \mathbb{E} \frac{\|\gamma(m_1; f)\|^r}{\|f\|^r} \leq \lim_{\|f\| \to \infty} \frac{\overline{K}^r}{\|f\|^r} \mathbb{E}\|m_1\|^r = 0.$$

Taking the test function $\|x\|^r$ with (Cline and Pu, 1999, Theorem 3.2 and their expression (3.4)) finalizes the proof as $\lim_{\|f\|\to\infty} \sup \|\theta(f)\| = \lim_{\|f\|\to\infty} \sup \|\frac{\omega+\beta f}{1+\|f\|}\| \leq \|\beta\| < 1$ by the eigenvalue condition on β. Beta-mixing is then implied by geometric ergodicity (see also Francq and Zakoian, 2011, p.66).

We show the existence of second moments to establish covariance stationarity of the strictly stationary backwards limit ${}_u f_t$ of $f_t(f_0)$ by a direct calculation. Let $\Omega := \mathbb{E} m_1 m_1^\top$.

$$
\begin{aligned}
\|\mathbb{V}\mathrm{ar}[{}_u f_t]\| &= \left\|\mathbb{V}\mathrm{ar}[\lim_{t\to\infty} f_t(f_0)]\right\| = \left\|\mathbb{V}\mathrm{ar}\left[\lim_{t\to\infty} \sum_{i=0}^{t} \beta^i \alpha K(f_{t-i})\, m_{t-i}\right]\right\| \\
&= \Big\| \lim_{t\to\infty} \sum_{i=0}^{t} \beta^i \alpha \mathbb{E}\left[f_{t-i} K(f_{t-i})^\top \mathbb{E}_{t-i+1}[m_{t-i} m_{t-i}^\top] K(f_{t-i})^\top\right] \alpha^\top (\beta^\top)^i \\
&\quad + \sum_{i<j} \mathbb{E}\left[\beta^i \alpha K(f_{t-i}) m_{t-i} \mathbb{E}_{t-i}\left[m_{t-j}^\top K(f_{t-j})^\top \alpha^\top (\beta^\top)^j\right]\right] \\
&\quad + \sum_{i>j} \mathbb{E}\left[\mathbb{E}_{t-j}\left[\beta^i \alpha K(f_{t-i}) m_{t-i}\right] m_{t-j}^\top K(f_{t-j})^\top \alpha^\top (\beta^\top)^j\right] \Big\| \\
&\leq \lim_{t\to\infty} \left\| \sum_{i=0}^{t} \beta^i \alpha \mathbb{E}\left[\sup_{f_{t-i}} \{K(f_{t-i})^\top \Omega K(f_{t-i})^\top\}\right] \alpha^\top (\beta^\top)^i \right\| \qquad (3.42)
\end{aligned}
$$

where we used Fubini's theorem, because expression (3.42) stays finite as $t \to \infty$ if β has modulus less than one, due to the boundedness conditions on $K(\cdot)$. □

Proof of Corollary 1. As the equicorrelation matrix depends only on one parameter, we can pick any decomposition $h(f_t)$ to derive the score. To simplify notation,

we drop the arguments of $\rho(f_t)$, $h(f_t;\lambda)$, and related functions and simply write ρ, h, etc. We also write $\dot{\rho}$ and $\ddot{\rho}$ for the first and second derivative of $\rho(f_t)$ with respect to f_t. A similar notation holds for the derivatives of other functions with respect to f_t. To derive the expressions for ρ_1 to ρ_6, define

$$\begin{aligned}
\rho_7 &= \sqrt{1-\rho}, && (3.43)\\
\rho_8 &= n^{-1}\cdot\left(\sqrt{1+(n-1)\cdot\rho}-\sqrt{1-\rho}\right), && (3.44)\\
\rho_9 &= \rho_7^{-1}=(1-\rho)^{-1/2}, && (3.45)\\
\rho_{10} &= n^{-1}\cdot\left((1+(n-1)\cdot\rho)^{-1/2}-(1-\rho)^{-1/2}\right). && (3.46)
\end{aligned}$$

We set

$$h=\rho_7\mathrm{I}_n+\rho_8\iota_n\iota_n^\top,\quad h^{-1}=\rho_9\,\mathrm{I}_n+\rho_{10}\iota_n\iota_n^\top, \tag{3.47}$$

such that $hh^\top = hh = h^2 = (1-\rho)\,\mathrm{I}_n + \rho\iota_n\iota_n^\top$. Define the functions

$$\dot{\rho} = 2n^{-1}\cdot(1-\rho)\,(1+(n-1)\cdot\rho)\,, \tag{3.48}$$

$$\rho_1 = -n^{-1}\cdot(1+(n-1)\rho), \tag{3.49}$$

$$\rho_2 = n^{-1}, \tag{3.50}$$

$$\rho_3 = -n^{-1}(1+(n-1)\rho)(1-\rho)^{-1}, \tag{3.51}$$

$$\rho_4 = n^{-1}\left((1+(n-1)\rho)^{-1}+\rho(1-\rho)^{-1}\right), \tag{3.52}$$

$$\rho_5 = \rho_9^2 = (1-\rho)^{-1}, \tag{3.53}$$

$$\rho_6 = (2\rho_9+n\rho_{10})\,\rho_{10} = n^{-1}\left((1+(n-1)\cdot\rho)^{-1}-(1-\rho)^{-1}\right). \tag{3.54}$$

We calculate the components in Proposition 1 by

$$\begin{aligned}
\Psi &= \mathrm{vec}(h^{-1}\dot{h}) = \dot{\rho}\cdot\mathrm{vec}\left(\ (\rho_9\mathrm{I}_n+\rho_{10}\iota_n\iota_n^\top)(\dot{\rho}_7\mathrm{I}_n+\dot{\rho}_8\iota_n\iota_n^\top)\ \right) && (3.55)\\
&= \dot{\rho}\cdot\mathrm{vec}\left(\ \dot{\rho}_7\rho_9\mathrm{I}_n+(\dot{\rho}_7\rho_{10}+\dot{\rho}_8\rho_9+n\dot{\rho}_8\rho_{10})\iota_n\iota_n^\top\ \right)\\
&= \dot{\rho}\cdot\mathrm{vec}\left(\ \rho_5\mathrm{I}_n+(\dot{\rho}_7\rho_{10}+\dot{\rho}_8\rho_9+n\dot{\rho}_8\rho_{10})\iota_n\iota_n^\top\ \right) && (3.56)\\
&= \rho_1(f_t;\lambda)\mathrm{vec}(\mathrm{I}_n)+\rho_2(f_t;\lambda)\iota_{n^2},\\
\nabla_{u,t} &= \Psi^\top\mathrm{vec}\left(w_{u,t}u_tu_t^\top-\mathrm{I}_n\right) = \mathrm{trace}\left(\dot{h}h^{-1}\left(w_{u,t}u_tu_t^\top-\mathrm{I}_n\right)\right)\\
&= \rho_1\left(-u_t^\top\nabla_{p_{u,\lambda}}(u_t)-n\right)+\rho_2\left(-\iota^\top\nabla_{p_{u,\lambda}}(u_t)u_t^\top\iota-n\right). && (3.57)\\
u_t^\top u_t &= y_t^\top h^{-1}h^{-1}y_t = y_t^\top(\rho_9\mathrm{I}_n+\rho_{10}\iota_n\iota_n^\top)^2y_t = \rho_9^2y_t^\top y_t+(2\rho_9+n\,\rho_{10})\,\rho_{10}\cdot(y_t^\top\iota)^2\\
&= \rho_5y_t^\top y_t+\rho_6\cdot(y_t^\top\iota)^2 = \frac{y_t^\top y_t}{1-\rho}+\frac{(y_t^\top\iota)^2}{n}\left(\frac{1}{1+(n-1)\cdot\rho}-\frac{1}{1-\rho}\right). && (3.58)
\end{aligned}$$

Due to $-\nabla_{p_{u,\lambda}}(u_t)^\top u_t = w_u(u_t^\top u_t;\lambda)\cdot u_t^\top u_t$, the required score expressions (3.24) and (3.25) follow.

As a special case, consider Student's t density. The i.i.d. score is then given by $\nabla_{p_{u,\lambda}}(u_t) = -w_tu_t$, where

$$w_t = -\Big((\nu+n)/(\nu-2)\Big)/\left(1+\frac{u_t^\top u_t}{\nu-2}\right).$$

Finally, the Hessian can be computed directly from Lemma 1 in Creal et al. (2011)

and by using $(1/g)\mathcal{I}_{p_{u,\lambda}} = 2\mathcal{D}\mathcal{D}^+ + \mathrm{vec}(\mathrm{I}_n)\mathrm{vec}(\mathrm{I}_n)^\top$ we have

$$\begin{aligned}
\mathcal{I}_t(f_t) &= \Psi(f_t;\lambda)^\top \left[g\mathcal{I}_{p_{u,\lambda}} - \mathrm{vec}(\mathrm{I}_n)\mathrm{vec}(\mathrm{I}_n)^\top\right] \Psi(f_t;\lambda) \\
&= (\rho_1(f_t;\lambda)\mathrm{vec}(\mathrm{I}_n) + \rho_2(f_t;\lambda)\iota_{n^2})^\top \left[g\left(2\mathcal{D}\mathcal{D}^+ + \mathrm{vec}(\mathrm{I}_n)\mathrm{vec}(\mathrm{I}_n)^\top\right) - \mathrm{vec}(\mathrm{I}_n)\mathrm{vec}(\mathrm{I}_n)^\top\right] \\
&\qquad\qquad (\rho_1(f_t;\lambda)\mathrm{vec}(\mathrm{I}_n) + \rho_2(f_t;\lambda)\iota_{n^2}) \\
&= n\left[n(g-1)(\rho_1+\rho_2)^2 + 2g\left(\rho_1^2 + 2\rho_1\rho_2 + n\rho_2^2\right)\right],
\end{aligned}$$

where $g = (\nu+n)/(\nu+n+2)$. Compared to the available expressions in the score literature, this is an improvement as the information matrix expression can be computed at the cost of $O(1)$ matrix operations. In the two-dimensional case, we have $\mathcal{I}_t(f_t) = 1+\rho^2$ for the Gaussian density ($g=1$). □

Proof of Corollary 2. The result follows from the proof of Proposition 2, if we show that the predetermined factor $K(f_t;\lambda) = S(f_t;\lambda)\Psi(f_t;\lambda)^\top$ from (3.22) is bounded. For a matrix $A \in \mathbb{R}^{m\times n}$ of rank r we can make use of the inequalities $\|A\|_2 \le \|A\|_F \le \sqrt{r}\|A\|_2$; $\|A\|_F \le \|A\|_* \le \sqrt{r}\|A\|_F$; $\|A\|_{\max} \le \|A\|_2 \le \sqrt{mn}\|A\|_{\max}$; $\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty$; $\frac{1}{\sqrt{m}}\|A\|_1 \le \|A\|_2 \le \sqrt{n}\|A\|_1$; ;$\|A\|_2 \le \sqrt{\|A\|_1\|A\|_\infty}$; such that the choice of norm is without loss of generality.

We discuss each of the choices $a \in \{0, 1/2, 1\}$ below.

(i) Consider $a = 1/2$. The square root scaling is defined by $S(f_t;\lambda) = \mathcal{J}(f_t)^\top$ with

$\mathcal{I}_t(f_t)^{-1} = \mathcal{J}(f_t)^\top \mathcal{J}(f_t)$. Without loss of generality, we recall the spectral norm:

$$\|K(f_t;\lambda)\|_2 = \max \sqrt{\text{eig}\{K(f_t;\lambda)^\top K(f_t;\lambda)\}}.$$

In the case of a Gaussian weighing matrix, we have $\mathcal{H}_{p_{u,\lambda}} = \mathcal{D}\mathcal{D}^+$ and therefore

$$K(f_t;\lambda)^\top K(f_t;\lambda) = \Psi(f_t;\lambda)\left(\Psi(f_t;\lambda)^\top \Psi(f_t;\lambda)\right)^{-1}\Psi(f_t;\lambda)^\top$$

can be recognized the Moore-Penrose as an idempotent form with eigenvalues on the unit disc.

However if the weighing matrix $\mathcal{H}_{p_{u,\lambda}}$ is allowed to follow a general form, no closed form expression exists for μ due to the necessity to solve high-degree polynomials for the above expression. However, any eigenvalue μ, corresponding to eigenvector v of the matrix $K(f_t;\lambda)^\top K(f_t;\lambda)$ satisfies

$$\mu v = \Psi(f_t;\lambda)\left(\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda)\right)^{-1}\Psi(f_t;\lambda)^\top v.$$

As $\text{rank}(\Psi(f_t;\lambda)) = k < n$, we can pre-multiply both sides
by $v^\top \Psi(f_t;\lambda)(\Psi(f_t;\lambda)^\top \Psi(f_t;\lambda))^{-1}\Psi(f_t;\lambda)^\top$ to obtain

$$\mu v^\top \Psi(f_t;\lambda)(\Psi(f_t;\lambda)^\top \Psi(f_t;\lambda))^{-1}\Psi(f_t;\lambda)^\top v =$$
$$v^\top \Psi(f_t;\lambda)\left(\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda)\right)^{-1}\Psi(f_t;\lambda)^\top v.$$

Positive definiteness of $h(f_t)$ implies full rank of $\Psi(f_t;\lambda)$ and $\text{rank}(\mathcal{H}_{p_{u,\lambda}}) \geq k$, if $\Psi(f_t;\lambda)^\top v = 0$, this implies $\mu = 0$ (as $v \neq 0$ by definition). Non-zero eigenvalues can then be characterized by

$$\mu = \frac{v^\top \Psi(f_t;\lambda)\left(\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda)\right)^{-1} \Psi(f_t;\lambda)^\top v}{v^\top \Psi(f_t;\lambda)(\Psi(f_t;\lambda)^\top \Psi(f_t;\lambda))^{-1}\Psi(f_t;\lambda)^\top v} \tag{3.59}$$

$$\leq \sup_{\tilde{v}\in\mathbb{R}^k} \frac{\tilde{v}^\top \left(\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda)\right)^{-1} \tilde{v}}{\tilde{v}(\Psi(f_t;\lambda)^\top \Psi(f_t;\lambda))^{-1}\tilde{v}} \tag{3.60}$$

And by the Rayleigh-Ritz theorem, we have

$$\mu \leq \frac{\max \text{eig}\left\{\left(\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda)\right)^{-1}\right\}}{\min \text{eig}\{(\Psi(f_t;\lambda)^\top \Psi(f_t;\lambda))^{-1}\}} = \sup_{\theta, f_t} \frac{\min \text{eig}\{\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda)\}}{\max \text{eig}\{\Psi(f_t;\lambda)^\top \Psi(f_t;\lambda)\}} < \infty \tag{3.61}$$

due to the existence of the inverses in the numerator and the denominator for all f_t.

(ii) Consider $a = 0$. Without loss of generality,observe for the Frobenius norm that

$$\begin{aligned}
\sup_f \|\Psi(f_t;\lambda)\|_F &= \sup_f \sqrt{\sum_{i=1}^{n^2}\sum_{l=1}^{k} |\{(\mathrm{I}_n \otimes h(f_t;\lambda)^{-1})\, \partial \text{vec}(h(f_t;\lambda))/\partial f_t^\top\}_{i,l}|^2} \\
&\leq \sup_f \sqrt{\sum_{i=1}^{n^2}\sum_{l=1}^{k} |\{(\iota_n \iota_n^\top \otimes h(f_t;\lambda)^{-1})\, \partial \text{vec}(h(f_t;\lambda))/\partial f_t^\top\}_{i,l}|^2} \\
&\leq n^4 k \max_{i,j,r,s} \sup_f |\partial h_{r,s}(f_t;\lambda)/\partial f_l|.
\end{aligned}$$

(iii) Consider $a = 1$. In addition to the calculations in *(ii)* we have

$$\sup_f \left\| \left(\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda) \right)^{-1} \right\|_2 = \sup_f \max \text{eig} \left\{ \left(\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda) \right)^{-1} \right\}$$
$$= \sup_f \min \frac{1}{\text{eig} \left\{ \left(\Psi(f_t;\lambda)^\top \mathcal{H}_{p_{u,\lambda}} \Psi(f_t;\lambda) \right) \right\}} < \infty,$$

where the last inequality follows from the full rank condition.

□

Remark 1. A higher order GAS(p, q) model of the form $f_{t+1} = \omega + \sum_{i=1}^{p} \beta_i f_{t+1-i} + \sum_{j=1}^{q} \alpha_j s_{t+1-j}$, where $s_t = K(f_t;\lambda) m_t$ can embedded in the reasoning of Proposition 2 by reformulating the stochastic difference equation as $\mathbf{f}_{t+1} = \mathbf{w} + \mathbb{B}\mathbf{f}_t + \mathbb{A}\mathbb{K}(\mathbf{f}_t;\lambda)\mathbf{m}_t$, where $\mathbf{f}_t = (f_t^\top, f_{t-1}^\top, \ldots, f_{t-p+1}^\top)^\top$,$\mathbf{m}_t = (m_t^\top, m_{t-1}^\top, \ldots, m_{t+1-q}^\top)^\top$, $\mathbf{w} = (\omega^\top, 0^\top, \ldots, 0^\top)^\top$ and

$$\mathbb{B} = \begin{pmatrix} \beta_1 & \beta_2 & \ldots & \beta_p \\ I_k & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & I_k & 0 \end{pmatrix} \qquad \mathbb{A} = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_q \\ 0 & \ldots & \ldots & 0 \\ \vdots & & & \vdots \\ 0 & \ldots & \ldots & 0 \end{pmatrix}$$

$$\mathbb{K}(\mathbf{f}_t;\lambda) = \begin{pmatrix} K(f_t;\lambda) & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & K(f_{t+1-q};\lambda) \end{pmatrix}. \tag{3.62}$$

The structure of the companion form is therefore conform to Proposition 2.

Proof of Corollary 3. We directly verify

$$\left|h(f_t;\lambda)^{-1}_{i,j}\partial h_{r,s}(f_t;\lambda)/\partial f\right| \leq C\left|(1-\rho)\sqrt{1+(n-1)\rho}-(1+(n-1)\rho)\sqrt{1-\rho}\right| \leq Cn.$$

□

Proof of Proposition 3. Consider the stationary solution and denote the L^r-norm by $\|\cdot\|_r$, we then have by successive applications of the Minkowski inequalities

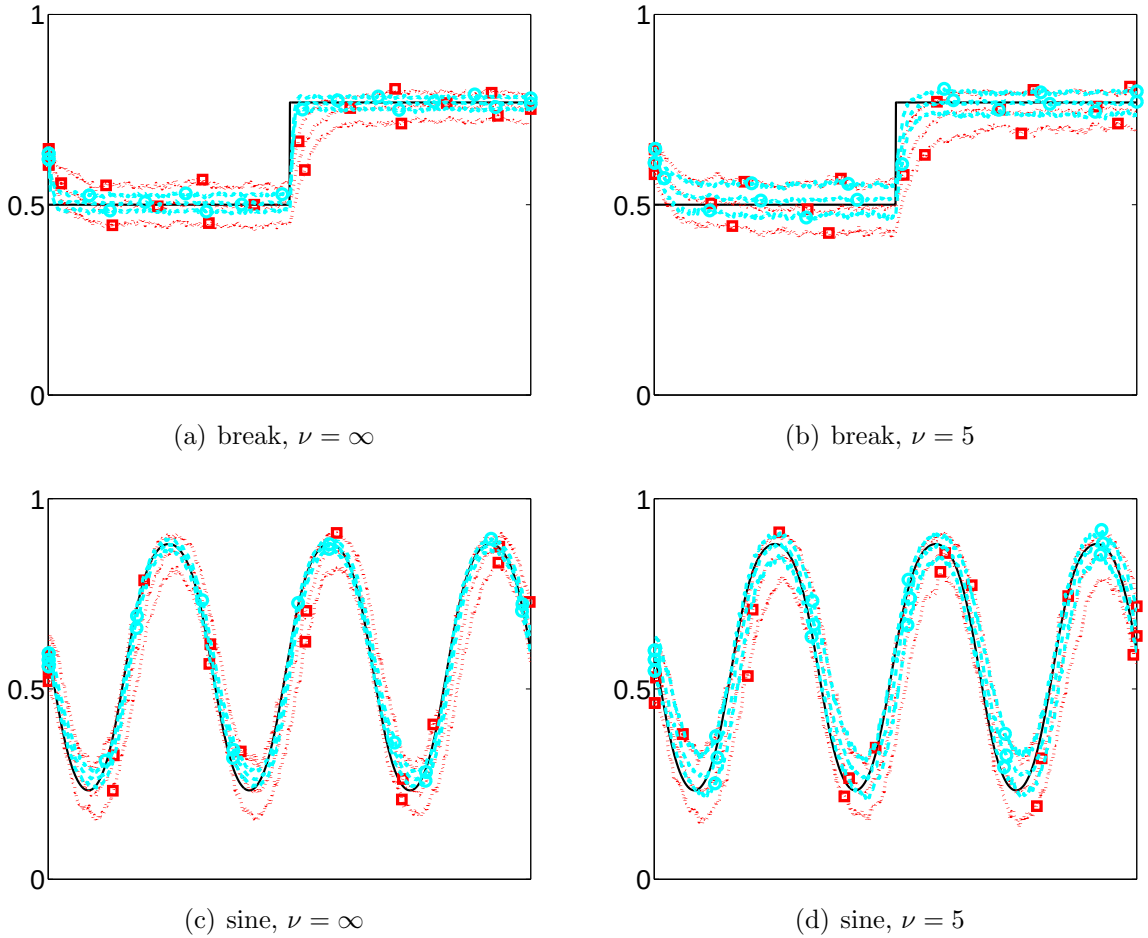
$$\begin{aligned}\|{}_u f_t\|_r &= \left\|(\mathrm{I}_n-\beta)^{-1}\omega+\sum_{i=0}^{\infty}\beta^i\alpha s_{y,t-i}\right\|_r \\ &\leq \left\|(\mathrm{I}_n-\beta)^{-1}\omega\right\|+\sum_{i=0}^{\infty}\mathbb{E}\left[\left\|\beta^i\alpha s_{y,t-i}\right\|^r\right]^{\frac{1}{r}}.\end{aligned}$$

We can further approximate

$$\sum_{i=0}^{\infty}\mathbb{E}\left[\left\|\beta^i\alpha s_{y,t-i}\right\|^r\right]^{\frac{1}{r}} \leq \sum_{i=0}^{\infty}\|\beta\|^i\|\alpha\|\,\overline{K}\,\|m_1\|_r<\infty,$$

due to the eigenvalue condition on β and thus $\|\beta\|<1$ as well as the condition $\|m_1\|_r<\infty$. □

3.B Supplemental information

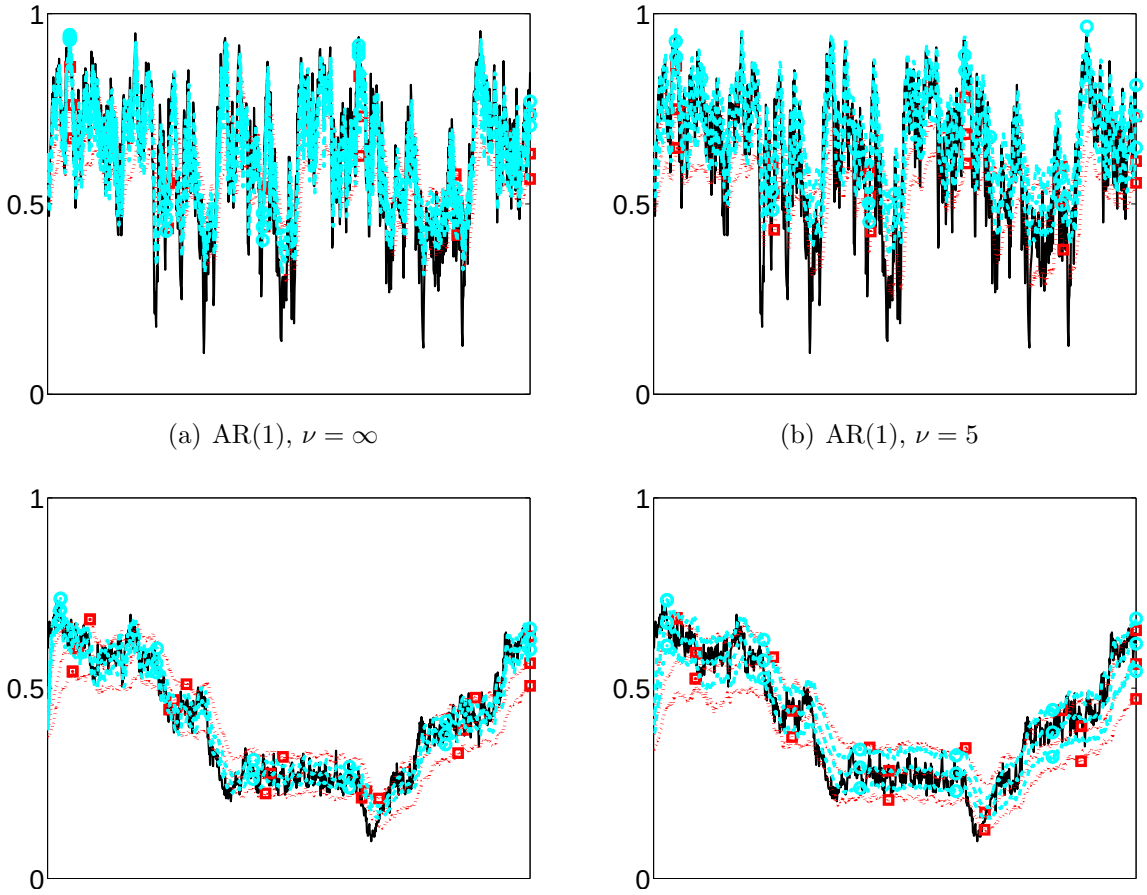


(a) break, $\nu = \infty$ (b) break, $\nu = 5$

(c) sine, $\nu = \infty$ (d) sine, $\nu = 5$

Figure 3.4: Signal extraction under deterministic misspecification

Notes: The dark solid line presents the true time-varying equicorrelation pattern; the DCC-DECO model is presented by red squares $(\cdots\square\cdots)$; the GAS-DECO model is presented by blue circles $(-\circ-)$. We present the pointwise average of the filtered correlations based on the QML estimates of θ for cross-sectional dimension $n = 100$ and 100 replications. The dotted bands depict quantiles (10% and 90%) of time-varying correlation estimates across simulations.



(a) AR(1), $\nu = \infty$

(b) AR(1), $\nu = 5$

(c) Local Level (LL), $\nu = \infty$

(d) Local Level (LL), $\nu = 5$

Figure 3.4: Signal extraction under deterministic misspecification (continued)

Notes: The dark solid line presents the true time-varying equicorrelation pattern; the DCC-DECO model is presented by red squares ($\cdots\square\cdots$); the GAS-DECO model is presented by blue circles ($-\circ-$). We present the pointwise average of the filtered correlations based on the QML estimates of θ for cross-sectional dimension $n = 100$ and 100 replications. The dotted bands depict 80% pointwise confidence-bands across simulations.

Table 3.5: The full universe of stocks included in the empirical analysis

Ticker	Company Name	Ticker	Company Name	Ticker	Company Name	Ticker	Company Name
A	Agilent Technologies Inc	CSX	CSX Corp	JPM	JP Morgan Chase & Co	ROP	Roper Industries Inc
AA	Alcoa Inc	CVC	Cablevision Systems Co A	JWN	Nordstrom Inc	SCG	SCANA Corp
AAPL	Apple Inc.	CVX	Chevron Corp	K	Kellogg Co	SCHW	Schwab Charles Corp
ABT	Abbott Laboratories	DD	E. I. du Pont de Nemours and Company	KO	Coca-Cola Co	SIAL	Sigma-Aldrich Corp
AES	AES Corp	DELL	Dell Inc	KR	Kroger Co	SLB	Schlumberger Ltd
AET	Aetna Inc	DHI	Horton D.R. Inc	LEG	Leggett & Platt	SLM	SLM Corp
AIG	American Intl Group Inc	DIS	Walt Disney Co	LEN	Lennar Corp	SNDK	SanDisk Corp
AIV	Apartment Investment & Mgmt	DOV	Dover Corp	LLY	Lilly Eli & Co	STJ	St Jude Medical Inc
AKAM	Akamai Technologies Inc	DUK	Duke Energy Corp	LM	Legg Mason Inc	SYMC	Symantec Corp
ALL	Allstate Corp	DVN	Devon Energy Corp	LMT	Lockheed Martin	SYY	Sysco Corp
ALTR	Altera Corp	EBAY	eBay Inc.	LOW	Lowe's Cos Inc	T	AT&T Inc
AMAT	Applied Materials Inc	ED	Consolidated Edison Inc	LUV	Southwest Airlines Co	TAP	Molson Coors Brewing Co B
AMD	Advanced Micro Devices	EIX	Edison Intl	MAS	Masco Corp	TE	TECO Energy Inc
AMGN	Amgen Inc	EOG	EOG Resources	MCD	McDonald's Corp	TEG	Integrys Energy Group Inc
AON	Aon plc	ETFC	E*TRADE Financial Corp	MHP	McGraw-Hill Cos Inc	TER	Teradyne Inc
APA	Apache Corp	EXC	Exelon Corp	MKC	McCormick & Co	TGT	Target Corp
APD	Air Products & Chemicals Inc	F	Ford Motor Co	MMC	Marsh & McLennan Companies	TWX	Time Warner Inc
AVY	Avery Dennison Corp	FAST	Fastenal Co	MMM	3M Co	TYC	Tyco Intl
AXP	American Express Co	FCX	Freeport McMoRan Copper & Gold	MO	Altria Group Inc	UNP	Union Pacific Corp
BA	Boeing Co	FE	FirstEnergy Corp	MRK	Merck & Co Inc	USB	US Bancorp
BAC	Bank of America Corp	FRX	Forest Laboratories	MS	Morgan Stanley	UTX	United Technologies Corp
BAX	Baxter Intl Inc	GE	General Electric Co	MSFT	Microsoft Corp	VMC	Vulcan Materials Co
BBY	Best Buy Co Inc	GPS	Gap Inc	MSI	Motorola Solutions Inc	VZ	Verizon Communications Inc
BF-B	Brown-Forman Corp B	GT	Goodyear Tire & Rubber Co	MTB	M&T Bank Corp	WAG	Walgreen Co
BLL	Ball Corp	HCBK	Hudson City Bancorp	MU	Micron Technology Inc	WFC	Wells Fargo & Co
BMY	Bristol-Myers Squibb	HCP	HCP Inc	MYL	Mylan Inc.	WFM	Whole Foods Market Inc
BRCM	Broadcom Corp A	HD	Home Depot Inc	NBR	Nabors Industries Ltd	WHR	Whirlpool Corp
BRK-B	Berkshire Hathaway B	HON	Honeywell Intl Inc	NEE	NextEra Energy Inc	WMT	Wal-Mart Stores
C	Citigroup Inc	HPQ	Hewlett-Packard Co	NUE	Nucor Corp	XL	XL Group Plc
CAH	Cardinal Health Inc	HRB	Block H & R Inc	NVDA	Nvidia Corp	XOM	Exxon Mobil Corp
CAT	Caterpillar Inc	HSY	Hershey Foods Corp	ORCL	Oracle Corp	XRAY	Dentsply Intl
CCE	Coca-Cola Enterprises	HUM	Humana Inc	PBI	Pitney Bowes Inc	XRX	Xerox Corp
CCI	Crown Castle Intl Corp	IBM	Intl Business Machines Corp	PCAR	PACCAR Inc	YHOO	Yahoo Inc
CCL	Carnival Corp	IGT	Intl Game Technology	PFE	Pfizer Inc	YUM	Yum! Brands Inc
CERN	Cerner Corp	INTC	Intel Corp	PG	Procter & Gamble		
CLX	Clorox Co	IPG	Interpublic Group Cos	PLD	ProLogis Inc		
CMI	Cummins Inc	IRM	Iron Mountain Inc	POM	Pepco Holdings Inc		
COP	ConocoPhillips	JDSU	JDS Uniphase Corp	PPL	PPL Corp		
COST	Costco Wholesale Corp	JNJ	Johnson & Johnson	RF	Regions Financial Corp		

Table 3.6: Stocks included in the option index analysis on May 29, 2009

Ticker	Company Name	Weight	Ticker	Company Name	Weight
AAPL	Apple Inc.	0.0292	MCD	McDonald's Corp	0.0158
ABT	Abbott Laboratories	0.0168	MDT	Medtronic Inc	0.0093
AMGN	Amgen Inc	0.0125	MMM	3M Co	0.0096
BAC	Bank of America Corp	0.0174	MO	Altria Group Inc	0.0085
BMY	Bristol-Myers Squibb	0.0095	MON	Monsanto Co.	0.0109
CMCSA	Comcast Corp A	0.0096	MRK	Merck & Co Inc	0.014
COP	ConocoPhillips	0.0164	MSFT	Microsoft Corp	0.0385
CSCO	Cisco Systems Inc	0.0261	ORCL	Oracle Corp	0.0181
CVS	CVS Caremark Corp.	0.0105	OXY	Occidental Petroleum	0.0131
CVX	Chevron Corp	0.0322	PEP	PepsiCo Inc	0.0195
DIS	Walt Disney Co	0.0108	PFE	Pfizer Inc	0.0247
GE	General Electric Co	0.0343	PG	Procter & Gamble	0.0367
GILD	Gilead Sciences Inc	0.0095	PM	Philip Morris International	0.0206
G	Google Inc	0.0241	QCOM	QUALCOMM Inc	0.0173
GS	Goldman Sachs Group Inc	0.0175	SGP	Schering-Plough Corp	0.0096
HD	Home Depot Inc	0.0095	SLB	Schlumberger Ltd	0.0165
HPQ	Hewlett-Packard Co	0.0199	T	AT&T Inc	0.0352
IBM	Intl Business Machines Corp	0.0344	UPS	United Parcel Service Inc B	0.0123
INTC	Intel Corp	0.0211	USB	US Bancorp	0.0081
JNJ	Johnson & Johnson	0.0368	UTX	United Technologies Corp	0.012
JPM	JP Morgan Chase & Co	0.0335	VZ	Verizon Communications Inc	0.0201
KFT	Kraft Foods Inc A	0.0093	WFC	Wells Fargo & Co	0.0261
KO	Coca-Cola Co	0.0236	WMT	Wal-Mart Stores	0.0268
LLY	Lilly, Eli & Co	0.0084	WYE	Wyeth	0.0144
LOW	Lowe's Cos Inc	0.0067	XOM	Exxon Mobil Corp	0.0827

Notes: The basket weights are determined by market value.

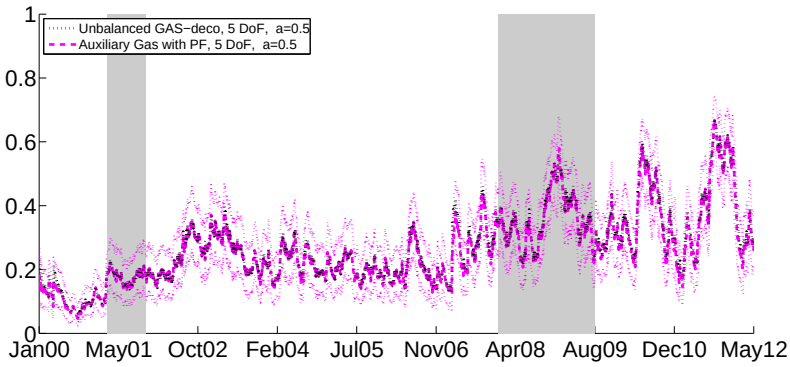


Figure 3.5: Comparison of GAS with particle filter output

Notes: The particle filter output corresponds to the parameter driven model with i.i.d. normal innovations, where the dynamic parameters θ are estimated via QML, cf. Table 3.2. The particle filter is implemented using the sequential importance sampling with resampling (SISR) method. The dotted lines represent the 90% confidence intervals around the filtered correlations.

Chapter 4

Adapting Filtering Equations to Misspecification

Abstract. This chapter adapts time-varying parameter models to misspecification. We measure misspecification by empirical deviations from model-implied moment conditions on time-varying parameters and observed data. It can be empirically tested whether these moment conditions hold for any particular sample at hand. If implied moment conditions fail to hold, alternative models are proposed with partially dynamic hyperparameters. We provide further generalizations of the standard GAS models in two examples. First, we allow for the intercept parameter to slowly vary over time rather than being a static hyperparameter. Second, we embed leverage effects in a filtering equation for volatility.

4.1 Introduction

While the previous chapters have dealt with dynamic correlation models, modern time series econometrics not only concerns modeling dynamic parameters, but also accounting for changing *patterns* in latent processes. For instance, one pattern that may change over time is the persistence of a latent process. The asymptotic analysis of the models provided in previous chapters of this thesis suggests that as more data become available, the estimated hyperparameters converge to some well-defined (pseudo-) true values. However, in empirical settings we may find that the supposedly constant pseudo-true parameters can vary substantially when we change the estimation sample. In this chapter we present a filtering framework that accounts for such dynamic misspecification by introducing time variation in dynamic hyperparameters of a misspecified time-varying parameter model. We explicitly relate new, revised, or extended model parameterizations to failed diagnostic checks for their simplified counterparts. As in the rest of the thesis we restrict our attention to observation driven models, which produce observable proxies for latent state variables.

It is generally straightforward to derive a set of model implied moment conditions that directly follow from the model specification. Furthermore, in the context of observation driven models, the data-based counterparts of model-implied moment conditions are explicitly observed. Our approach centers around proposing alternative or extended model parameterizations that better account for local deviations from these moment conditions.

For instance, consider the phenomenon of the Great Moderation stemming from macroeconomics, which is associated with the period of lower volatility in output due to business cycle fluctuations as of the mid 1980s until the Great Recession. When applying off-the-shelf models, fitting models in different sub-samples can lead to different implications for the fundamental characteristics of the model such as the long-run level, persistence and signal-to-noise ratio – the so-called "deep parameters" of the model, i.e., the parameters that fundamentally characterize the system settings and behavior. This is a particularly fundamental issue within the economics discipline, as we might evaluate policy interventions based on such fundamental data patterns.

We propose models with time-varying hyperparameters that better accommodate misspecification of a dynamic model with constant hyper-parameters. An alternative approach would be to fit time-varying parameter models on a rolling window basis. However, in the context of dynamic non-linear models no closed-form expressions are generally available for the maximum likelihood estimator. Therefore, a rolling window procedure would require choosing the rolling window length and numerical likelihood maximization within each of the rolling windows. This computationally cumbersome approach is circumvented in our model-based approach.

As our first example, we revisit the models from previous chapters by particularizing the analysis to the familiar dynamic correlation models. We consider dynamic models that allow the long-run mean to evolve over time. In the context of linear processes, Dahlhaus et al. (1997) proposed so-called "locally stationary" models that

capture stationary deviations from a non-stationary mean. Such a model structure may prove useful if the data displays more serial dependence than expected from a standard non-stationary model. Diebold (1986) even argues that omitted structural breaks may spuriously generate long memory. There is extant literature available for modeling shifts in the local mean of time-varying parameter models.[1] Most closely related to the example of changing long term means, Bauwens et al. (2013) accommodate changing levels in a trivariate DCC correlation model by a multistep procedure.

As our second example, we consider modeling time-varying volatility. We extend robust GARCH models to include a leverage-type term in the model. By explicitly combining the asymmetry and robustness features of the model, the model outperforms other available volatility models in terms of diagnostic checks and out-of-sample forecasting performance.

The structure for the rest of this paper is as follows. Section 4.2 introduces the concepts to guide us through the rest of the chapter. Section 4.3 introduces a time-varying GAS model to capture variation in latent hyperparameters of the filtering models. We find that constant dynamic hyperparameter specifications are rejected in

[1] Engle and Lee (1999) propose a GARCH unobserved components model that incorporates time-varying volatility levels. Furthermore, taking the route of local spectral representations, Stărică and Granger (2005) argue that it is sufficient to model stock market volatility by piecewise constant paths. Dahlhaus et al. (2006) develop the theory for locally stationary ARCH processes; see also Dahlhaus (2012) for a survey. Amado and Teräsvirta (2013) develop a specification test for time-variation in unconditional volatility levels and subsequently adopt smooth threshold autoregressive (STAR) type models for capturing this source of non-linearity. Koo and Linton (2012) semi-parametrically extract latent time-variation in the continuous time diffusion context. Giraitis et al. (2014) semi-parametrically extract time-varying parameters under a high persistence assumption, prominently providing confidence bands for impulse response functions in a vector autoregressive framework.

the empirical example of modeling correlations between UK and Greek stock indices. By using a locally stationary version of the GAS model, one can achieve performance gains in terms of out-of-sample model fit. Section 4.4 discusses the misspecification of unconditional moment restrictions in the context of omitted leverage in robust volatility filters. Finally, Section 4.5 concludes.

4.2 Adapting observation driven filters to misspecification

Consider data y_t characterized by the conditional density $y_t \sim p(y_t|f_t, \mathcal{F}_{t-1})$, where $\mathscr{F}_t$ denotes the filtration generated by the history of observations that are available at time t and f_t is a true univariate time varying parameter, such as for instance a time varying mean, volatility, or correlation. We filter the data y_t for an estimate of the unknown time-varying parameter. All observation driven models in this thesis satisfy the following decomposition of the filtering equation

$$f_{t+1} = \gamma\left(f_t, \theta\right) + \Delta\left(y_t; f_t, \theta\right), \tag{4.1}$$

where θ denotes a vector of static hyperparameters. We assume the measurability properties $\gamma_t := \gamma\left(f_t, \theta\right) \in \mathscr{F}_{t-1}$, $\Delta_t := \Delta\left(y_t; f_t, \theta\right) \in \mathscr{F}_t$. The interpretation is that γ_t captures the predictable drift term while Δ_t describes the observation driven

innovation term. For instance, in the specific case of a correctly specified plain vanilla GAS models with square root information matrix scaling, we have $\gamma_t = \omega(1-\beta)+\beta f_t$, $\Delta_t = \alpha \mathcal{I}_t^{-1/2} \nabla(y_t; f_t)$ and $\theta = (\omega, \alpha, \beta)$, where $\nabla(y_t; f_t)$ denotes the score of the predictive conditional density and $\mathcal{I}_t$ denotes the time $t-1$ Fisher information matrix.

For a data generating process from the GAS model, if the conditional density $p(y_t|f_t, \mathcal{F}_{t-1})$ and the transition dynamics are correctly specified, then for the model-implied true parameters $\theta = \theta_0$ and correct initialization of f_0, we have the restriction $\mathbb{E}_{\theta_0}\left[\Delta\left(y_t; f_t, \theta_0\right) | \mathscr{F}_{t-1}\right] \equiv 0$, where $\mathbb{E}_{\theta_0}\left[\cdot | \mathscr{F}_{t-1}\right]$ denotes the model expectation under the true parameter θ_0. Therefore, the moment restriction for the conditional mean is of the form

$$\mathbb{E}_{\theta_0}\left[f_{t+1} | \mathscr{F}_{t-1}\right] = \gamma\left(f_t^0, \theta_0\right). \tag{4.2}$$

In practice, the data may not conform to the above moment restriction for the conditional mean. We may then have to come up with an improved functional form of the filtering equation that would alleviate misspecification in the sense of the moment restriction violation. For example, for the plain vanilla GAS model with $\gamma(f_t, \theta) = \omega + \beta f_t$ we may find that over repeated observations the realized f_t exceeds the conditional mean $\gamma(f_t, \theta)$. If such deviations are persistent, the information in the data might be better exploited by providing an alternative model parameterization with a conditional mean $\tilde{\gamma}(f_t, \theta)$ that moves more closely in line with the realizations, e.g., that moves upwards if the realized time-varying parameters persistently exceed the conditional mean. This can be achieved my making the originally static

hyperparameter ω time varying.

The main task of the econometrician is now to come up with sensible transition equations for f_{t+1} that are more flexible and give fewer violations of model implied moment conditions. The extent of improvement in fit following from such proposals is an empirical issue, which we test in an out-of-sample analysis and by virtue of the diagnostic tests that were used to design the new adapted filtering models. Concrete examples of proposed new transition equations for f_{t+1} that allow for deviations from moment conditions are provided in Section 4.3 and in Section 4.4. Such multi-level modeling requires the data to be informative enough to identify variation in the violated moment conditions.

The above approach for time variation in the conditional mean of the time-varying parameter can be generalized further. Consider a set of K moment conditions generated under the correct specification of the recurrence equations (4.1),

$$m_t\left(\{y_s, f_s\}_s; \theta_0\right) := \mathbb{E}_{\theta_0}\left[g_t(\{y_s, f_s\}_s) \,|\, \mathscr{F}_{t-1}\right],$$

where $g_t(\cdot)$ and $m_t(\cdot)$ are known functions of the observed data and time-varying parameter paths. In the above example of time-varying conditional mean, the choices reduce to $g_t(\{y_s, f_s\}_s) := f_{t+1}$ and $m_t(\{y_s, f_s\}_t) := \gamma(f_t, \theta_0)$. If this moment condition is persistently violated in sample, we can make some of the (originally static) hyperparameters dynamic to exploit better the information contained in the data. This

can be done by proposing an alternative filtering equation that has a better match to its implied moment conditions.

For example, in the leading example of specifying the condition mean of the GAS model, we have a new functional form

$$f_{t+1} := \omega_t(1-\beta) + \beta f_t + \alpha \nabla_t.$$

The above parametrization serves the purpose of injecting omitted information into the filtering recursion by making ω_t a dynamic parameter. Two concrete examples are provided in the next two subsections.

4.3 The tvGAS model

We extend the standard score driven model (see also Creal et al., 2013; Harvey, 2013) to allow for time-variation in latent moments. We propose the following updating recursion,

$$\begin{aligned} y_t &\sim p(y_t; f_t) \\ f_{t+1} &= (1-\beta)\omega_t + \beta f_t + \alpha \mathcal{I}_t^{-a} \nabla(y_t; f_t), \qquad a \in \{0, 1/2, 1\} \end{aligned}$$

which we call the tvGAS model given the time-varying intercept parameter ω_t. In an observation driven framework the hyperparameters are modeled by $\omega_t = \omega_t(y_1, \ldots, y_{t-1}) \in$

$\mathscr{F}_{t-1}$. In order to interpret the role of ω_t, assume for now that the intercept parameter $\omega_t = \overline{\omega}$ would be constant and that the process f_t is invertible. The filter can be written as

$$f_{t+1} = \overline{\omega} + \sum_{i=0}^{\infty} \beta^i \alpha \mathcal{I}_{t-i}^{-a} \nabla(y_{t-i}; f_{t-i}), \tag{4.3}$$

and $\omega_t = \overline{\omega}$ can be interpreted as the long-run mean of f_t under correct specification. On the other hand, if ω_t is allowed to vary over time, the representation no longer continues to hold. In that sense there is no "stationarity" concept available here as the model itself changes over time. For that reason, the literature refers to "local stationarity". Furthermore, note that the extended model with flexible intercept term is no longer Markovian.

In empirical data we might observe long-lasting persistent deviations from the long-term means. In order to capture slow changes in the long-term time-varying

parameter levels, we consider the following flexible specification of long-run levels[2]

$$\omega_t = K_t^{-1} \sum_{s=1}^{T} k_s^*(t) f_s \tag{4.4}$$

$$k_s^*(t) = \begin{cases} k_s(t) & \text{if } s \leq t, \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

$$K_t = \sum_{s=1}^{T} k_s^*(t). \tag{4.6}$$

We follow Giraitis et al. (2014), suggesting the kernel

$$k_s(t) := K\left(\frac{t-s}{H}\right), \tag{4.7}$$

where $K(\cdot)$ denotes a standardized kernel, such as the Gaussian kernel with $K(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi}$. Theoretical results by Giraitis et al. (2014) suggest an optimal bandwidth choice of $H = \sqrt{T}$.[3] [4] These choices for weights are illustrated in Figure 4.1 using the Gaussian kernel.

[2]We have defined the model to have only non-zero weights on innovations from past observations. Alternatively, one can develop a smoothing algorithm by incorporating information from adjacent future levels in the filtering equation (4.6) in a two-sided sum.

[3]An alternative approach of Bauwens et al. (2013) defines the kernel by $k_s(t) := K_h\left((t-s)/T\right)$, where $K_h(\cdot) = (1/h)K(\cdot/h)$. They use optimal bandwidths of $h = 0.05$ for the Gaussian kernel and $h = 0.09$ for the Epanechikov kernel. Furthermore, by following the non-parametric approach, the method of (Hardle, 1990, p.127) can provide the time series confidence bands around time-varying parameters.

[4]Note that as opposed to extracting the optimal bandwidth semi-parametrically, the model here specifies a kernel-driven evolution of the latent deep state parameters.

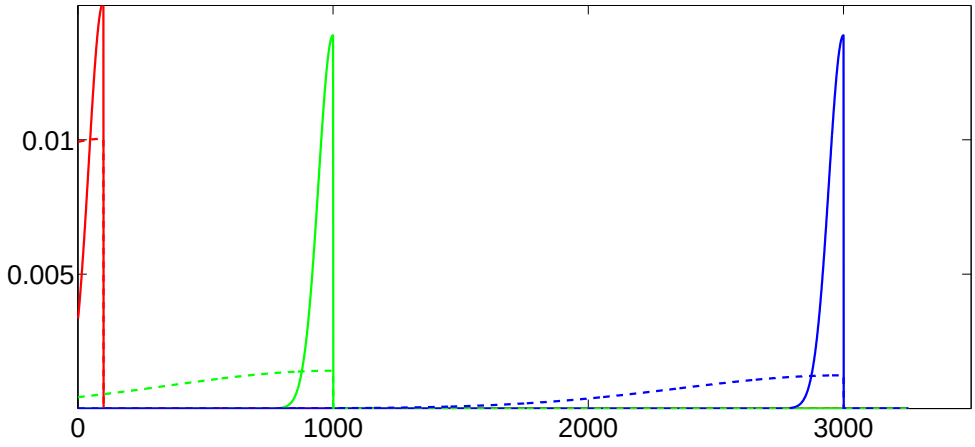


Figure 4.1: Kernel weights as a function of time

Notes: The thick dotted line corresponds to the bandwidth $H = T^{4/5}$, while the thin solid line corresponds to the optimal bandwidth of $H = \sqrt{T}$. The base kernel is Gaussian.

4.3.1 Monte Carlo evidence

The design of our simulation exercise is close to the analysis conducted by Engle (2002a). We simulate data with deterministic time-varying correlations and estimate our tvGAS model on this data in order to see how well our model can pick up the simulated patterns. In particular, we might be concerned about the additional parameter uncertainty of the tvGAS model overshadowing the benefits of its added flexibility.

To explore the tradeoff between estimation uncertainty and mis-specification bias, we parametrize the correlation process as $\rho_t = \tanh(f_t)$, where $f_{t+1} = \omega_t(1-\beta) + 0.99\, f_t + 0.01\, \varepsilon_t$, where $\varepsilon_t \sim NID(0,1)$. Furthermore, the observations are assumed

to be drawn from a Student's $t(5)$ distribution. We consider two types of changes in the long-run correlation levels: single breaks and sinusoidal patterns. Within each pattern we also vary the magnitude of the structural change to make it more difficult to extract. In the first class of a single break, we consider both a large break and a small break. In the second class of simusoidal waves we consider the case of a single period and four periods. By fixing the time series length to $T = 3000$, we reflect an empirical setting to capture over a decade of daily data. In order to capture changes of parameter uncertainty in the observation equation, we also vary the cross-sectional dimension $n \in \{2, 100\}$ and report the results over 1000 repeated simulations. Both GAS and tvGAS specifications are initialized at the true value of the time-varying parameter.

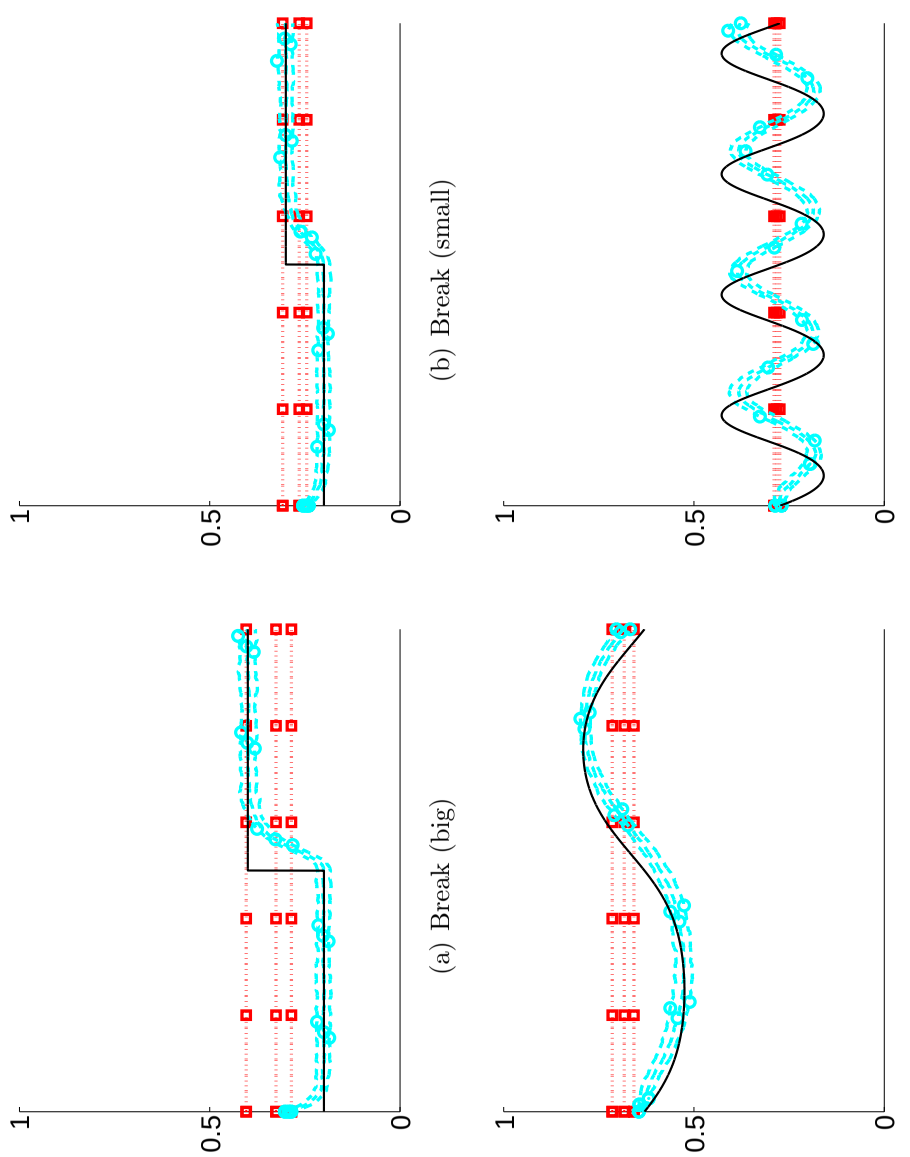


Figure 4.2: Steady state (ω_t) extraction under deterministic misspecification

Notes: The dark solid line presents the true time-varying steady state pattern; the GAS model is presented by red squares ($\cdots\square\cdots$); the tvGAS model is presented by light blue circles ($-\circ-$). We present the pointwise average of the filtered correlations based on the ML estimates for time-dimension $T = 3000$ and cross-sectional dimension $n = 100$, based on 1000 replications. The dotted bands depict 80% pointwise confidence-bands across simulations.

Table 4.1 summarizes the results. We are interested in the signal extraction quality both in terms of the correlation paths and the inferred steady states. We also display the true and the extracted steady-state parameter paths in Figure 4.2. In terms of the effective correlation paths, the average mean squared error (MSE) and median absolute deviation (MAD) of the tvGAS model are found to be fairly similar to the original GAS model. However, if we want to estimate the long run steady state correlation as a structural parameter, the gains in terms of the estimated long run steady state can be large. Furthermore, when we increase the dimensionality and therefore the information from cross-sectional pooling, we also observe that the signal extraction quality of both models improves.

We find that the tvGAS model performs best in case of a large single structural break in the middle of the sample. A smaller magnitude of the structural break or frequent substantial structural changes in the underlying steady state make the true data generating more difficult to extract. In more noisy scenarios, performance of the tvGAS model deteriorates. We put the tvGAS model to the test in such a hostile environment by also including a high-frequency sinusoidal specification of the long run steady state. Even though the tvGAS performs relatively better in a higher dimensional (informationally richer) setting, we do observe that such a process marks the breaking point of our model – both the correlation paths and the steady state extraction performance of the tvGAS specification are inferior to the plain vanilla GAS counterparts. Economically speaking, we would however not expect the

Table 4.1: Signal extraction results

Notes: L_x denotes the average loss $L \in \{MSE, MAD\}$ with respect to the respective time-varying parameter $x \in \{\rho, \omega\}$. The steady state paths of the simulated data generating process are displayed in Figure 4.2. In particular, "Break (big)" corresponds to a DGP where the steady state correlation switches from 0.2 to 0.4 at midpoint, while "Break (small)" corresponds to a DGP where the steady state correlation switches from 0.2 to 0.3. Furthermore, "Sine (slow)" represents a sinusoidal pattern with a single period, while "Sine (fast)" represents a sinusoidal pattern with four periods.

(a) $n = 2$

		MSE_ρ	MAD_ρ	MSE_ω	MAD_ω
Break (big)	GAS	0.0726	0.0458	0.1252	0.0718
	tvGAS	0.0669	0.0441	0.0725	0.0384
Break (small)	GAS	0.0568	0.0375	0.0616	0.0417
	tvGAS	0.0540	0.0371	0.0501	0.0341
Sine (slow)	GAS	0.0727	0.0484	0.2259	0.1717
	tvGAS	0.0736	0.0484	0.1107	0.0840
Sine (fast)	GAS	0.0666	0.0498	0.0974	0.0954
	tvGAS	0.0798	0.0610	0.1155	0.0982

(b) $n = 100$

		MSE_ρ	MAD_ρ	MSE_ω	MAD_ω
Break (big)	GAS	0.0288	0.0159	0.1150	0.0746
	tvGAS	0.0250	0.0158	0.0439	0.0113
Break (small)	GAS	0.0225	0.0141	0.0585	0.0363
	tvGAS	0.0214	0.0138	0.0254	0.0101
Sine (slow)	GAS	0.0231	0.0151	0.1081	0.1003
	tvGAS	0.0230	0.0151	0.0341	0.0259
Sine (fast)	GAS	0.0268	0.0181	0.0957	0.0954
	tvGAS	0.0271	0.0181	0.0949	0.0857

long-run steady-state to change abruptly from year to year as the average business cycle is expected to last around five years (see also Diebold and Rudebusch, 1994; Stock and Watson, 2005).

4.3.2 Illustration: Bivariate correlation between UK and Greek equity indices

For our first empirical illustration, we use data of the FTSE 100 and the Athex Composite index returns to study the time varying correlation between these two markets. We estimate the model over the period January 1, 2002 to December 31, 2011. To avoid overfitting, we study the performance from an out of sample perspective, from January 2, 2012 until June 24, 2014.

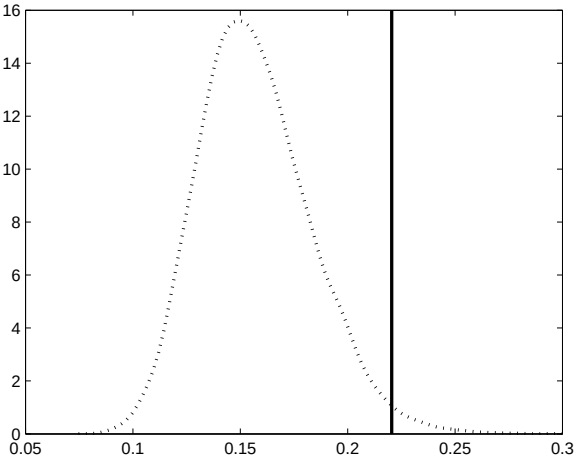


Figure 4.3: The estimated long run factor variance compared to its filtered counterpart

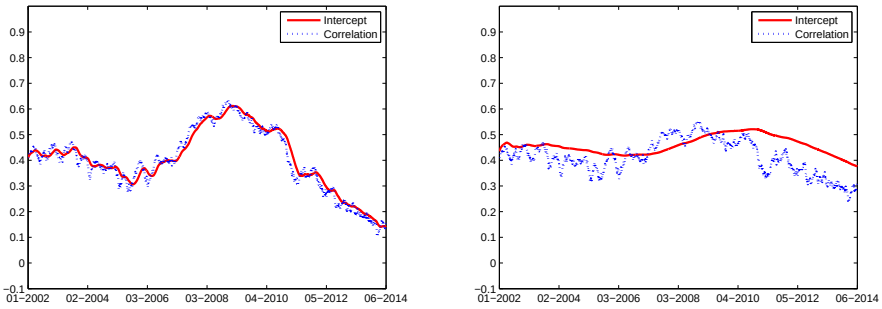
Notes: The vertical line depicts the Newey-West long-run variance of the filtered factors $f_t(y_t;\widehat{\theta})$. We compare this value to its simulated counterpart $f_t(\widehat{\theta})$, where we have simulated from the fitted model in Table 4.2. We find that the realized long-run variance lies on the 98.46%-quantile of the simulated distribution.

We have already seen in Chapter 2 an indication of possible higher order depen-

dence in extracted correlations on the UK-Greek equity index data. The Nyblom test for remaining time varying correlations in the residuals of the t(5)-GAS(1,1) model suggests that a further component to model correlations might be beneficial. As a further diagnostic check we compare variation in the model-implied correlation factors against the corresponding data counterpart. First, we compute the Newey-West long-run variance of filtered factors $\{f_t(y_t;\widehat{\theta})\}$.[5] Next, given the estimated model parameters $\widehat{\theta}$, we repeatedly simulate time varying parameter paths $f_t(\widehat{\theta})$ from to the the standard GAS correlation model. We do this for the fixed sample size to reflect the finite estimation sample size. Subsequently, we calculate the corresponding Newey-West long-run variance in each simulation just as before. This procedure finds that the realized long-run variance is located on the simulated 98.46%-quantile of the simulated long-run variance distribution, reported in Figure 4.3. Therefore, in our modeling approach, we explore excess variation above and beyond a time-varying volatilities and correlation t(5)-GAS(1,1) model. According to Engle's ARCH test, there does not seem to be substantial heteroskedasticity in the scores, we proceed by specifying a tvGAS model with time-varying long-run levels.

Giraitis et al. (2014) suggest the optimal bandwidth choice in (4.7) to be $H = \sqrt{T}$. We visualize the impact of bandwidth choice in Figure 4.4. We observe that under the alternative bandwidth choice $H = T^{4/5}$, the extracted evolution of the latent level ω_t

[5]Given a time-varying parameter path $\{f_t(y_t;\widehat{\theta})\}_{t=1,\dots,T}$, the heteroskedasticity and autocorrelation consistent long-run variance of the time-varying parameter proxy is computed by using lag length $4 \times (T/100)^{\frac{2}{9}}$ (see e.g. Davidson and MacKinnon, 2004).



(a) $H = \sqrt{T}$ (b) $H = T^{4/5}$

Figure 4.4: The impact of kernel choice on extracted time-varying correlation paths for fixed dynamic hyperparameters

Notes: We have used the kernel definition of (Giraitis et al., 2014), where the base kernel is Gaussian.

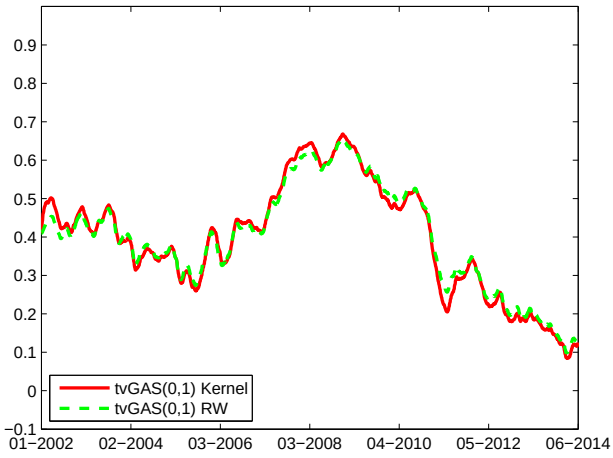


Figure 4.5: Comparison of kernel-based law of motion to the random walk intercept evolution in (4.8)

is much slower, effectively inducing patterns comparable to a plain vanilla GAS(1,1) model.

Also, we compare the kernel-based law of motion with the stochastic process

intercept assumption

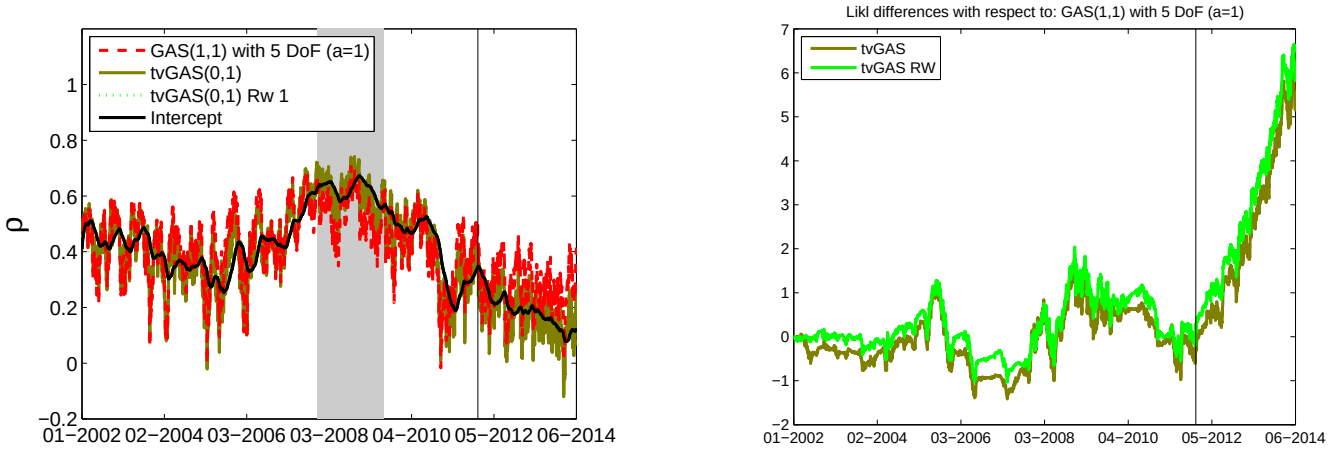
$$\omega_t = \omega_{t-1} + \alpha_\omega \left(f_{t-1} - \omega_{t-1} \right). \tag{4.8}$$

This random walk specification is labeled as the "tvGAS RW" model. Figure 4.5 shows that the extracted time-varying parameter paths of both specifications are fairly similar, such that this modeling choice does not substantially affect qualitative implications of the tvGAS specification in our example.

Table 4.2: Performance comparison in the bivariate case (5 DoF, a=1)

Notes: "tvGAS Kernel" denotes a model for the intercept process $\{\omega_t\}$ with kernel-extracted evolution (4.4). "tvGAS RW" denotes a model for the intercept process $\{\omega_t\}$ with a stochastic process parametrization (4.8). "Giacomini-White statistic" corresponds to the Giacomini and White (2006) test for superior conditional out of sample predictive ability, using GAS(1,1) model as the benchmark (critical value at 5% level: 5.99).

	GAS(1,1)	tvGAS	tvGAS RW
ω	0.0126		
α	0.0320	0.0289	0.0299
β	0.9734	0.9493	0.9544
α_ω			0.0148
Log-likelihood In-sample	−7004	−7005	−7004
Log-likelihood OOS	−1835	−1829	−1829
AIC (Full-sample)	17686	17673	17674
BIC (Full-sample)	17704	17685	17692
# parameters	3	2	3
Reject Const Corr at 5% level	Yes	Yes	No
Giacomini-White statistic	0	8.28	8.19



(a) Filtered correlations

(b) Differences in cumulative likelihoods

Figure 4.6: Comparison with a plain vanilla t(5)-GAS(1,1) model

Notes: The solid vertical line marks the beginning of the out-of-sample environment from January 2, 2012 until June 24, 2014. The smooth black solid line denotes the time-varying attractor ω_t, extracted by a semi-parametric Gaussian kernel with bandwidth $H = \sqrt{T}$.

Table 4.2 reports the fitted parameter estimates and various measures for model performance. We notice that the estimated persistence parameter β falls below 0.95. The estimated persistence parameter being close to unity (i.e. approaching the boundary of the admissible parameter-space) could be attributed to mis-specification (Diebold, 1986; Jensen and Lange, 2010). Therefore, according to this interpretation, mis-specification is alleviated.[6] Furthermore, a more precise estimate of the currently prevailing local time-varying parameter level requires less severe corrections by the score in each updating step because the model formulation embeds the additional information from recent correlation levels. This has the effect that the value of α is smaller for the new tvGAS model. In addition, we see that the GAS model extracts higher correlations in the early sample (from January 2002 to March 2006), lower correlations during the Great Recession (from December 2007 to June 2009) and higher correlations out-of-sample(from January 2012 to June 2014), when compared to the tvGAS models. Panel (b) of Figure 4.6 demonstrates the relevance of such a model compared to a benchmark t(5)-GAS correlation model, especially in an out-of sample setting. The out-of sample period differs from the in-sample data due to the European Securities Market Programme, injecting over EUR 200 billion into bond markets in June 2012, which we expect to have an impact on the whole financial ecosystem. Even though this information is not explicitly available to the new tvGAS

[6]It is important to avoid simply overfitting a more complicated model to the data. This can be achieved either by means of economic intuition or statistical methodology. To address the latter approach, we conduct out-of-sample analyses and diagnostic tests, verifying that the improvements are of substance.

model in the in-sample fitting stage, it quickly adapts to the post-crisis circumstances to deliver superior forecasts when compared to the plain vanilla GAS(1,1) model.[7]

4.4 Asymmetry in the volatility news impact curve

While the above example was concerned with dynamic violations of latent conditional moment conditions, more generally, one could also involve unconditional moment conditions that are explicit functions of the data. An example for such a mis-specification is the omission of the leverage effect by the GARCH model, i.e. the observation that negative returns are associated with higher future volatility than positive returns.

As a starting point for our exercise, consider a dynamic scale model based on the Student's t distribution with ν degrees of freedom

$$y_t = \exp\left(\frac{1}{2} f_t\right) u_t, \qquad u_t \overset{\text{i.i.d.}}{\sim} T_\nu\left(0, 1\right)$$

One can directly verify that the Student's t score, being a function of y_t^2, is symmetric in y_t. Therefore, this plain vanilla GAS(1,1) model generates the following symmetry

[7]The tvGAS also prevails in a full-sample model fitting procedure.

restriction

$$\mathrm{E}_{\theta_0}\left[\log\left(y_{t+1}^2\right)\left(\mathbf{1}\left[y_t < 0\right] - \mathbf{1}\left[y_t \geq 0\right]\right)\right] = 0. \tag{4.9}$$

The logarithmic transformation of the data serves three purposes. First, the logarithmic transformation of the volatility proxies brings them close to Gaussianity (Andersen et al., 2007), which makes standard regression-based diagnostic checking more reliable. Second, we formulate restrictions directly in terms of observable data rather than in terms of the latent volatility processes. Potentially, measurement error in volatility can convolute the estimated persistence and therefore alter the extracted time-varying parameter paths and diagnostics thereof (see also Hansen and Lunde, 2014). Third, given the exponential parametrization of volatility as a function of the time-varying parameter, a logarithmic transformation of the realized volatility proxy yields an analytically convenient link between the model and the moment restriction violations. Engle and Ng (1993) have argued that the above symmetry restriction is likely to be violated due to the leverage effect in equity markets (Black, 1976). Typically, diagnostic tests yield a higher impact of negative returns to expected volatility, i.e. the expectation in (4.9) is positive rather than zero for typical empirical data.

If the data indeed disagrees with the symmetry restriction, we may want to improve upon a symmetric volatility model. We can model deviations from the empirical

expectation the following mis-specification adapted model

$$y_t = \exp\left(\frac{1}{2} f_t\right) u_t, \qquad u_t \overset{\text{i.i.d.}}{\sim} T_\nu(0,1)$$

$$f_{t+1} = \omega(1-\beta) + \beta f_t + (\alpha + \alpha_- \mathbf{1}[y_t < 0]) \nabla_t(y_t; f_t), \tag{4.10}$$

which we label as the T-GAS Leverage model. We can directly compute the moment property (4.9) for the T-GAS Leverage model by

$$\mathbb{E}_\theta\left[\log\left(y_{t+1}^2\right)(\mathbf{1}[y_t < 0] - \mathbf{1}[y_t \geq 0])\right] = \alpha_- \, \mathbb{E}_\theta\left[\nabla_t(y_t; f_t) \,|\, y_t < 0\right] \tag{4.11}$$

to observe that the asymmetry parameter α_- in the new transformed recursion (4.10) captures the moment violation up to a constant of proportionality. In other words, the T-GAS Leverage model helps to improve upon the misspecification of the original model in the sense of the moment condition (4.9). Furthermore, by the nature of the explicit connection of the above moment condition to the newly introduced parameter α_- retains a tractable interpretation as the asymmetry parameter.

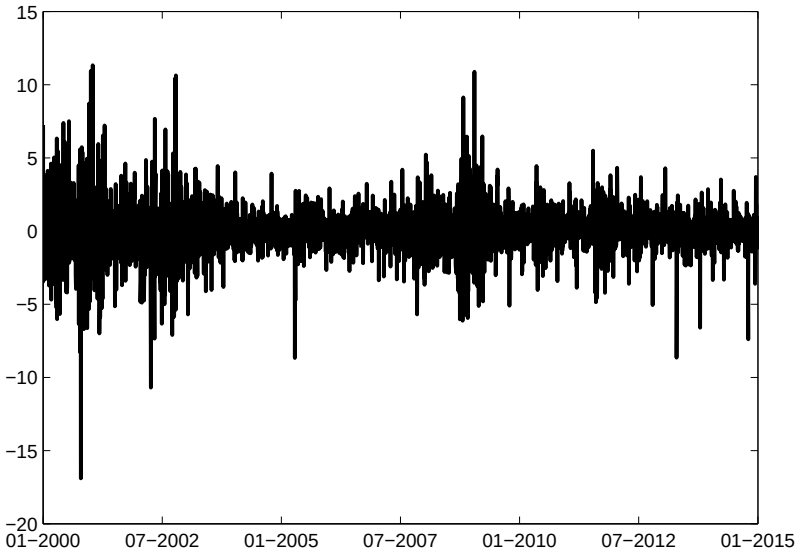
While the score of the predictive likelihood still continues to act as the key component of the updating recursion, we have further incorporated the salient asymmetry feature of stock market volatility by adapting the functional form of the volatility dynamics. Consequently, we prevent the volatility model from being entirely tied down to the observation density. Furthermore, this example illustrates how the choice of the relevant moment conditions can play a pivotal role in our inference

(Gallant and Tauchen, 1996). Note that the testable moment condition (4.9) is formulated in terms of a discontinuous function of the data. Therefore, the adjusted updating recursion inherits the discontinuity around zero. This complicates numerical computations and the asymptotic analysis of the new model, which may not be considered desirable by various researchers.

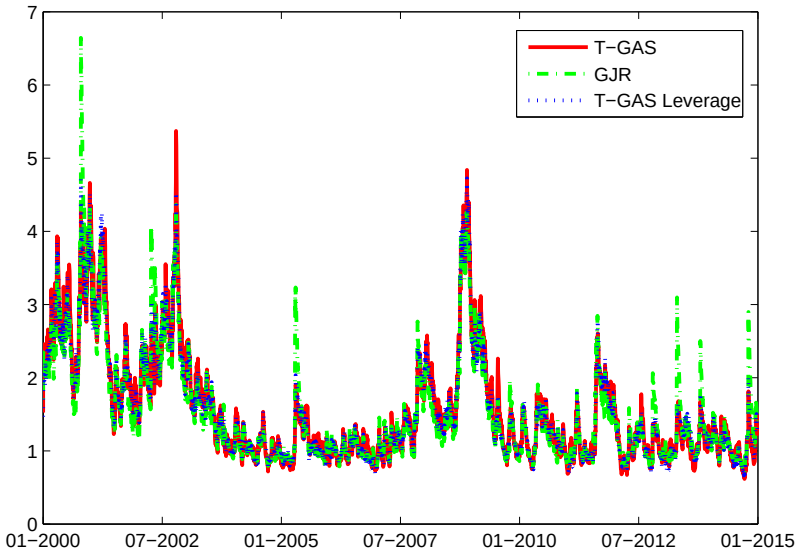
We illustrate the importance of this modeling choice based on IBM returns of between January 1, 2000 and December 31, 2014. Table 4.3 reports different performance criteria, while Figure 4.7 depicts the extracted volatility paths. We see that the log-likelihood of a GJR model of Glosten et al. (1993) with a fat-tailed observation density is on par with the corresponding values of GAS models based on Students T and skewed T densities.[8]

It may seem from the above comparison of existing volatility models that there is little benefit in adding asymmetry features to the volatility filters. However, our T-GAS Leverage model enables us to have a closer look. When comparing the leverage-adjusted T-GAS model to a benchmark GJR-GARCH model, we see that as the latter uses squared returns as inputs, large outliers come with single spikes in extracted volatilities. On the other hand, the transformed T-GAS Leverage recursion outperforms a plain vanilla T-GAS model by incorporating information from the above moment condition. For instance, in late 2002, after the burst of the tech bubble, the IBM stock return was positive and large in absolute value. While the plain

[8]The skewed T distribution is implemented by restricting the Generalized Hyperbolic $GH(\mu, \gamma, -\nu/2, \nu, 0)$, where γ denotes the skewness parameter.



(a) Returns



(b) Volatilities

Figure 4.7: IBM volatility models

vanilla TGAS and the TGAS Leverage specifications remain robust to this outlier, the GJR GARCH model overreacts due to its quadratic news impact curve. Further-

more, we also fit a volatility model based on a skewed T distribution, which allows for asymmetry in the news impact curve only if there is skewness in the observation density. However, in this setting as the skewness parameter is insignificant such that the observation density is estimated to be symmetric, and therefore the extracted paths and performance criteria are comparable to a symmetric Student's t GAS specification. Also note that our leverage specification, derived by (4.11), critically makes use of the symmetry assumption of the observation density.

In order to disentangle the fat tails, skewness and leverage channels in the news impact curve, we next turn to diagnostic tests. Table 4.3 also reports the sign statistic of Engle and Ng (1993) as a measure of asymmetry in the magnitude of filtered residuals. We observe a slight improvement of the TGAS Leverage model over the TGAS specification in terms of diagnostic performance. Summing up, the T-GAS Leverage specification outperforms other models both in terms of density forecasting and in terms of model specification tests.

Table 4.3: Volatility models and leverage

	T-GAS	SkewT-GAS	GJR	T-GAS Leverage
ω	0.839	0.837	0.026	0.764
	(0.245)	(0.244)	(0.006)	(0.210)
β	0.990	0.990	0.917	0.9903
	(0.004)	(0.004)	(0.009)	(0.004)
α	0.067	0.067	0.033	0.032
	(0.008)	(0.008)	(0.009)	(0.010)
α_-			0.079	0.050
			(0.015)	(0.011)
ν	5.0647	5.063	5.329	5.056
	(0.3805)	(0.353)	(0.406)	(0.375)
Skewness		0.004		
		(0.025)		
LogL	−6473	−6473	−6477	−6464
AIC	12954	12956	12963	12938
BIC	12979	12987	12994	12969
numParam	4	5	5	5
Sign statistic	6.263	6.209	1.628	4.372
p-value	0.099	0.102	0.653	0.224

Notes: The LM-type sign statistic of Engle and Ng (1993) is implemented by a regression

$$\log\left(\frac{y_t}{\sigma_t}\right)^2 = \gamma_0 + \gamma_1 \mathbf{1}\left[y_{t-1} < 0\right] + \gamma_2 \mathbf{1}\left[y_{t-1} < 0\right] y_{t-1} + \gamma_3 \mathbf{1}\left[y_{t-1} \geq 0\right] y_{t-1}$$

to test the restriction $\gamma_1 = \gamma_2 = \gamma_3 = 0$.

4.5 Conclusion

We have extended available filtering models that have fallen short in describing the data. In particular, if a model has failed some of the diagnostic checks, we explicitly incorporate the missing features into a new, better specified model. The applications have demonstrated usefulness of the approach, especially for forecasting purposes.

There remain further extensions that we have not yet examined. First, we have not developed asymptotic analysis of the novel filtering models. While in the linear GARCH case, explicit results are available in the literature, we suspect that for general non-linear models, high-level assumptions may need to be maintained. Second, we have sidestepped issues of overidentification to account for several contemporaneous sources of mis-specification. Third, we have focused on the filtering problem. Extensions to analogous smoothing algorithms can as well be constructed. Forward-looking data descriptions could be accommodated by applying two-sided temporal kernels in place of the one-sided kernels used in this chapter. Finally, there is ample room for further applications. For example, the proposed filtering method could be used equally well to capture possible time-varying persistence in economic data.

Chapter 5

Predictive Systems and Dynamic Spillover Analysis

Abstract. We propose dynamic vector autoregressive (VAR) models for the purpose of measuring systemic risk. Following the spillover decomposition framework of Diebold and Yilmaz (2009), we compare our dynamic empirical models to the prevailing rolling window approach. Our results suggest that the new model has superior performance both in terms of the in-sample predictive likelihood and out-of-sample loss criteria. Furthermore, our approach has the feature that the incorporated realized variance measures allow for more precise models for second moments and allows us to pin down variance decompositions that underlie the spillover measurement framework more precisely. We illustrate our approach in the context of equity return spillovers across listed US companies and volatility spillovers across international stock market indices.

5.1 Introduction

The transmission of shocks is an object of empirical interest for understanding and measuring financial crises. If movements in a single securities market predict further ripple effects or uncertainty in other international exchanges, this can have implications for policy interventions, economic equilibrium models and portfolio choice (Ahern, 2014; Aït-Sahalia and Hurd, 2014; Kodres and Pritsker, 2002). In this chapter we measure time-varying heterogeneous spillovers in equity markets. We introduce new models, which, by matching salient features of stock market data, provide a more credible description of systemic risks when compared with available approaches in the literature.

Recently, research on the sources and the measurement of systemic risks has emerged rapidly, see Bisias et al. (2012) for an overview. Analyzing spill-over externalities has become increasingly relevant in the face of policy interventions that involve wide-scale asset purchases. By disentangling the firm-specific heterogeneity of contagion, this may enable us to identify systemically important sources of spillovers.

We base our analysis on Diebold and Yılmaz (2014) who propose a network interpretation of vector autoregression (VAR) based variance decompositions. Vector autoregressions have been used before for the analysis of systemic risks: for instance Billio et al. (2012) monitor unconditional Granger causalities in equity returns and Billio et al. (2012) investigate Granger causalities in sovereign credit default spreads. However, mostly static VAR-based models are applied in the literature, and in order

to pick up changes in systemic risks, previous papers have mainly relied upon a rolling window estimation of a static model. When studying contagion among financial assets, a key issue to address is the changing nature of the economic data generating process (Timmermann, 2008). For instance, Bekaert et al. (2014) find strong crisis-related contagion effects in global equity markets. As we will see, a rolling window approach to parameter instability comes with the caveat of high parameter uncertainty and suffers from poor out-of sample predictive performance.

An improved model can provide more accurate descriptions of the data. We depart from the widespread rolling window approach by implementing an autoregressive observation driven recursion for the time-varying parameters of the underlying VAR model, which implies a robustified exponential weighting of the innovations and lagged dependent variables. Exponential weighting (EWMA) of the innovation is expected to provide from a forecasting perspective optimal weights in the presence of small continuous structural instabilities (Pesaran, Pick, and Pranovich, 2013). Furthermore, we expect substantial changes in the extracted spillover indices, as the rolling window approach often exhibits discontinuities, while exponential weighting of the innovations in our model will more gradually detect the inferred changes in spillover patterns.

Conditional models in the literature mostly address a *contemporaneous* factor structure to capture asset commonality across different firms (see also Lucas et al., 2013). Such a methodology indicates systemic risks in instances where all assets move

more strongly together. On the other hand, we aim to capture 'domino-effects' that could be symptoms of fire-sales and downward spirals in asset prices, which display *lead-lag* relationships. While maintaining the possibility of a contemporaneous factor structure, we incorporate further heterogeneity into the approximating model for identifying the systemically important sources of spillovers. Therefore our proposed methodology addresses a different, yet complementary, characteristic of systemic risks and financial crises.

One benefit of our approach is that we are able to deal with unbalanced panels of data. Therefore we avoid temporal aggregation of the data to economically uninteresting and artificially low frequencies. Missing observations necessarily occur in our international dataset as banking holidays are not synchronized across different countries. Moreover, handling missing observations may become particularly relevant in face of financial crises, as corporate default events or asynchronous trading holidays may lead to missing observations or exits from the available dataset.

We illustrate the alternative spillover methodologies in two examples. First, we study return spillovers among US large cap stocks in a timeframe between January 2001 and July 2014. Our empirical results suggest that in times of crises, the relative value of our systemic risk index goes above and beyond the strength of the factor structure as implied by the system correlation. While the financial assets were systemically central during the times of the 2008 financial crisis, we find that the technology stocks are the source of spill-overs in the early 2000's to capture the burst

of the internet bubble. This illustrates how different economic sectors may play a key role in crisis propagation in different historical episodes. Our spillover methodology helps uncover these time-varying patterns.

Second, we study volatility spillovers across global equity markets in a timeframe between June 2002 and June 2015. We find that US equity shocks lead the global markets. Furthermore, developed open economies are characterized by both giving shocks to and receiving shocks from other markets. Emerging markets on the other hand display a strong idiosyncratic component, being less intertwined with global markets. The important distinction between both examples is the estimated smoothness of the spillover patterns. While a rolling regression approach exogenously fixes the window length, the filtering models estimate the speed of information accumulation. As volatility proxies are more persistent than returns, we also find the spillover measures to behave more smoothly in the volatility spillover example.

This paper is organized as follows. Section 5.2 introduces the concept of spillover networks and proposes the dynamic vector autoregressive model. Sections 5.3 and 5.4 discuss the forecasting performances and spillover implications of the proposed method compared to the existing benchmarks in the context of return and volatility models, respectively. Finally, Section 5.5 concludes.

5.2 Dynamic spillover models

5.2.1 Variance decomposition networks

Let y_t denote the n-dimensional asset universe under investigation and consider an MA(∞) representation

$$y_t = \sum_{i=0}^{\infty} A_h \varepsilon_{t-i}, \qquad \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \Sigma).$$

Let d_{ij}^H denote the fraction of asset i's H-step variance explained due to asset j. The generalized variance decomposition of Pesaran and Shin (1998) gives the following analytic expression

$$d_{ij}^H = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} \left(e_i^\top A_h \Sigma e_j\right)^2}{\sum_{h=0}^{H-1} e_i^\top A_h \Sigma A_h^\top e_i}, \qquad i, j = 1, \ldots, n, \tag{5.1}$$

where e_i denotes the i-th unit vector.[1] In practice we approximate the MA(∞)-representation by a finite order VAR(p) process. For instance, if a VAR(1) system is sufficient to describe the joint dynamics, then we have $A_h = \Phi^h$, where Φ is the autoregressive matrix. We now describe the Diebold and Yılmaz (2014) network interpretation of variance decompositions, to explore the inferred system connectedness

[1] This variance decomposition has the property that $\sum_j d_{ij}^H \neq 1$, so in practice we normalize d_{ij} over the columns j. An alternative variance decomposition is via the Cholesky decomposition, which delivers very similar relative systemic risk rankings (see also Diebold and Yılmaz, 2014; Klößner and Wagner, 2014).

properties.

5.2.2 Descriptive spillover measures

A key descriptive measure within the Diebold and Yılmaz (2014) framework is the spillover index

$$SI = \frac{\sum_{i=1}^{n} \text{Contribution To Others}_i}{\sum_{i=1}^{n} \text{Contribution Including Own}_i},$$

$$\text{Contribution To Others}_i = 100 \times \sum_{k \neq i} d_{ki}^H,$$

$$\text{Contribution Including Own}_i = 100 \times \sum_{k=1}^{n} d_{ki}^H,$$

$$\text{Contribution From Others}_i = 100 \times \sum_{k \neq i} d_{ik}^H.$$

Note that the spillover table has the property $\sum_{k=1}^{n} d_{ik}^H = 1$ as we have row-normalized the spill-over measures. Furthermore, to disentangle the directions of shocks, define the 'Net' connectedness

$$NET_i = \text{Contribution To Others}_i - \text{Contribution From Others}_i,$$

with the interpretation that a positive value for NET_i is an indication for leading assets in the asset universe. Next, we further study the heterogeneous nature of spillovers. In order to pin down the systemically important assets, define the spillover

centrality

$$C_{\text{spillover},i} = 100 \times \frac{\text{Contribution To Others}_i}{\sum_k \text{Contribution To Others}_k}.$$

While the above definitions are straightforward in a static setting, it is unclear how the variance decompositions would behave when used for describing different dynamic models. Furthermore, the importance of incorporating dynamics into the systemic risk descriptions cannot be overemphasized, or as Diebold and Yılmaz (2014) phrase it: *"Given this background of financial market evolution and turbulence, it seems unlikely that any single fixed-parameter model would apply over the entire sample."* As conclusions regarding the variance decomposition networks may crucially hinge upon the dynamic approximating model for the data, we next explore sufficiently general and flexible multivariate models to represent the data generating process. In particular, we postulate observation-driven dynamics for the underlying VAR parameters as a convenient modeling tool in such a high-dimensional and computationally demanding setting.

5.2.3 Multivariate location dynamics

Consider the following observation driven model for the asset returns

$$y_t = \Phi_t y_{t-1} + u_t, \qquad u_t \sim \mathrm{N}(0, \Sigma_t) \tag{5.2}$$

$$\Leftrightarrow y_t = X_t f_t + u_t, \qquad X_t = \mathrm{I}_n \otimes y_{t-1}^\top, \qquad f_t = \mathrm{vec}\left(\Phi_t^\top\right)$$

$$\Sigma_t = D_t R_t D_t, \qquad D_t = \mathrm{diag}\left(\{\sigma_{i,t}\}_{i=1,\ldots,n}\right)$$

$$f_{t+1} = \omega + \beta f_t + \alpha s_t \tag{5.3}$$

$$s_t = \left(X_t^\top \Sigma_t^{-1} X_t\right)^+ X_t^\top \Sigma_t^{-1} \mathrm{diag}\left(w(y_t)\right)\ (y_t - X_t f_t)$$

$$= \frac{1}{y_{t-1}^\top y_{t-1}} \left(\mathrm{diag}\left(w(y_t)\right) \otimes y_{t-1}\right) \left(y_t - \left(\mathrm{I}_n \otimes y_{t-1}^\top\right) f_t\right) \tag{5.4}$$

$$w(y_t) = \left\{ \left(\frac{\nu+3}{\nu}\right) \frac{\nu+1}{\nu - 2 + (u_{t,i}/\sigma_{t,i})^2} \right\}_{i=1,\ldots,n}, \tag{5.5}$$

where the innovation s_t is motivated by a Student's t score (Creal et al., 2013, 2014). Furthermore, the variance matrix Σ_t is decomposed into a correlation matrix R_t and a diagonal matrix of volatilities D_t as $\Sigma_t = D_t R_t D_t$. The model for the variance matrix Σ_t depends on the application at hand and data availability. Concrete examples are provided in Sections 5.3 and 5.4.

Notice that the score based updating recursion for the full system is in the Gaussian case equivalent to updating the dynamic slope parameters equation-by-equation.[2] By updating the dynamic VAR matrix in an equation-by-equation manner, we treat asynchronous outliers separately across different assets. However, in a full-system esti-

[2]Note that the equation-by-equation updating is due to the block-diagonal form of $\left(X_t^\top \Sigma_t^{-1} X_t\right)^+$.

mation under the multivariate Student's t distribution, the variance matrix Σ_t would enter the updating equations via the outlier-robust weighting function $w(y_t)$. As systemic risk applications typically require estimating medium- to high-dimensional systems to model financial data, we choose to sacrifice the potential informational gain of a full-information estimation against the benefits of robustness of a limited-information approach to outliers and model mis-specification.

We require the admittedly restrictive Gaussianity assumption in the observation equation (5.2) in order to maintain the closed form variance decomposition in (5.1). While variance decompositions for non-linear non-Gaussian systems can be accommodated by the generalized impulse response functions of Koop et al. (1996), the Gaussianity assumption substantially reduces the required numerical effort for evaluating spillover risks. Furthermore, the Gaussianity may facilitate further theoretical research on the estimation of the model to justify a Quasi Maximum Likelihood interpretation and on the validity of a two-step estimation procedure (see e.g. Rahbek and Pedersen, 2013).

For the purposes of gaining further intuition, consider the dynamic regression model that is inherent to estimating the vector autoregression in an equation-by-equation manner. In the special case of $X_t \equiv 1$, $w(y_t) \equiv 1$, $\omega \equiv 0$ and $\beta \equiv 1$ the EWMA recursion can be recovered. We estimate the model parameters by maximum likelihood. Note that in a n-dimensional system, the intercept parameter ω collects n^2 parameters which can become large quickly even for moderately sized ($n > 10$)

systems. Therefore, we reduce the required numerical effort of fitting the model to data by targeting the intercept parameters in ω of the updating recursion (5.3) by simple ordinary least squares estimation of a static VAR model as the first stage of a location targeting procedure, which greatly reduces the number of numerically optimized parameters. The signal and persistence parameters α and β are scalars in our specifications to enforce parsimony in our relatively large nonlinear model.

A benefit of the model is that by letting $w(y_t)$ in the above definitions follow the score driven mechanics of Creal et al. (2013, 2014); Harvey and Luati (2014) based on the Student's t distribution, the model inferred time-varying parameter paths are resilient against contamination from aberrant observations. In our applications on equity markets we find evidence for excess kurtosis and we ensure that the filter is robust in the sense of Calvet et al. (2014) by using a weighting for the time varying parameter dynamics based on Student's t(5) scores.

5.3 Return spillovers

5.3.1 Multivariate scale dynamics

The location model in (5.2) – (5.5) left the dynamics of conditional covariance matrices unspecified. Depending on the choice of the asset universe and the observation frequency, we may consider different models for higher moments. Time variation in conditional covariances is a well-known stylized fact in equity markets, so we proceed

with modeling higher moments as well.

Neglecting time-varying volatilities in a vector autoregression can have several undesirable effects: observations belonging to a high-volatility regime are overemphasized; spuriously generated time-varying parameters emerge (Cogley and Sargent, 2005; Sims, 2001); and there is an upward bias of the estimated contagion effects (Forbes and Rigobon, 2002). Therefore we account for conditional heteroskedasticity by employing volatility clustering models. Accounting for heteroskedasticity has the additional advantage that a precise covariance measure helps to pin down the evolution of the long-run spectrum, and therefore inference regarding variance decompositions is expected to become more reliable.

Recent advances in the volatility forecasting literature emphasize the benefits of exploiting intraday data to obtain more precise volatility estimates. We treat the latent covariance matrices as partially measured from daily returns and from high-frequency data, building upon robustified innovations from the scalar BEKK model of Engle and Kroner (1995b) and the multivariate high-frequency-based volatility (HEAVY) model of Shephard et al. (2012), respectively. We specify the transition dynamics by

$$
\begin{aligned}
\Sigma_{t+1} &= (1-\beta^{\sigma})\Sigma + \beta^{\sigma}\Sigma_t + \alpha_1^{\sigma} w_t^{\sigma} s_{t,1}^{\sigma} + \alpha_2^{\sigma} w_t^{\sigma} s_{t,2}^{\sigma} \\
s_{t,1}^{\sigma} &= u_t u_t^{\top} - \Sigma_t, \qquad s_{t,2}^{\sigma} = \gamma^{\sigma} RK_t - \Sigma_t \\
w_t^{\sigma} &= \frac{\nu + n}{\nu - 2 + u_t^{\top}\Sigma_t^{-1} u_t},
\end{aligned}
$$

where RK_t denotes the realized kernel at time t (see also Barndorff-Nielsen et al., 2008).[3] The role of γ^σ is to capture the omission of overnight returns in the realized volatility calculations and we therefore expect to obtain an estimate larger than unity. This is where our model slightly deviates from pre-existing volatility models with realized proxies. Unlike the pre-existing models, which entirely focus on volatility modeling, the object of immediate interest in this chapter is the dynamics in the return equation. As the high information content in the measurement equation for realized proxies would inadvertently dominate the less informative return equation, we only incorporate the return equation directly into the statistical loss function for estimating the model parameters. We further reduce the number of numerically optimized parameters by tying the long run volatility level to the time series average of the high-frequency proxy, i.e. $\widehat{\Sigma} = (1/T)\sum_{t=1}^{T} \gamma^\sigma RK_t$.

5.3.2 Statistical evaluation of alternative VAR models

As the economically interesting near-term variance decompositions depend on the conditional forecasting ability over a short horizon, we expect a more accurate dynamic model also to deliver more precise descriptions of the inferred sources of spillovers. We apply our return spillover model to the universe of major US companies

[3]The high-frequency data is extracted from the TAQ database of the Wharton Research Data Services (WRDS). I would like to thank Anne Opschoor for providing the realized kernel data.

listed in Table 5.9 of Appendix 5.A. The time frame of our study spans the period between January 2, 2001 and July 31, 2014.

A static VAR system has many parameters to be estimated. The problem is possibly even aggravated in our present dynamic setting. In our context of modeling international equity markets, we inherently quantify the existence and degree of return predictability (for an overview, see also Timmermann, 2008; Welch and Goyal, 2008). As a high degree of estimation uncertainty of many parameters may harm the out-of-sample fit, we also explore the potential benefits of regularized estimators as a robustness check.

In order to reduce estimation uncertainty in our model, we investigate the impact of penalized estimation of the steady-state ω. The steady state estimate of the system can be obtained from the corresponding static model. By further penalizing the static model, we are able to regularize the steady state estimates. Our main workhorse is the adaptive LASSO procedure of Zou (2006).[4] By downweighting the penalty attached to large coefficients, the adaptive LASSO approach has several advantages over a basic LASSO regression. First, the procedure has improved model selection oracle properties under milder conditions than a plain vanilla LASSO

[4]In the first targeting step of the estimation procedure, we replace the least squares estimation of the intercept with a LASSO estimation routine of the intercept. The adaptive LASSO estimator is defined as the minimizer of the following optimization problem

$$\min_{\beta_i} \sum_{t=1}^{T-1} \left(y_{t+1,i} - y_t^\top \beta_i\right) + \lambda \sum_{p=1}^{n} \frac{1}{\left(\widehat{\beta_{i,p}}\right)^2} \left|\beta_{i,p}\right|,$$

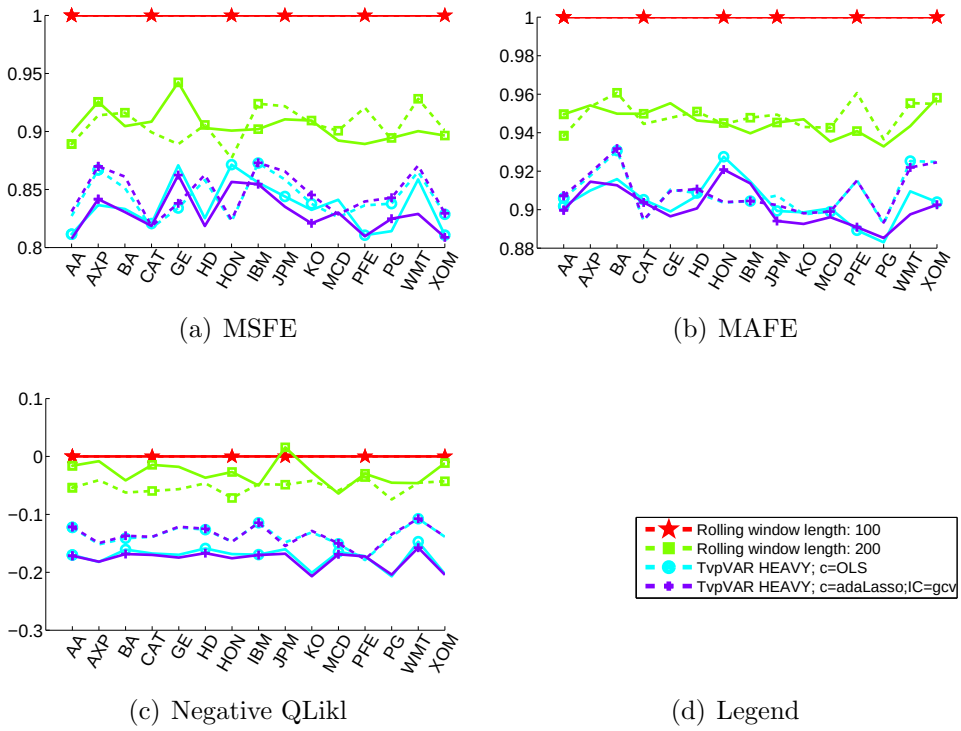
where $\widehat{\beta}_i$ is typically the OLS estimate corresponding to a constant coefficient model.

procedure. Second, the optimization problem is convex, and its global minimizer can be efficiently implemented.[5] Third, the adaptive LASSO has delivered solid performance across various empirical applications (see e.g. Callot and Kock, 2014; Hautsch, Schaumburg, and Schienle, 2014).

For the static VAR model, we adopt the approach of Diebold and Yilmaz (2009) to recursively estimate rolling window parameters based on only a subset of recent data and then subsequently forecast from the static model based on the current parameters. On the contrary, our dynamic model in equations (5.2) – (5.5) uses the entire in-sample data and the changes in the predictive dynamics are already built into the model via the updating recursion for the time-varying VAR parameters f_t.

In order to gain insight into the sources of performance discrepancies, we study the different models across the cross-sectional dimension and across time series dimension in Figures 5.1 and 5.2, respectively. We see that a short rolling window of 100 days (from Diebold and Yılmaz, 2014) has the best performance across all models in capturing the initial conditions. Moreover, the improved performance of the rolling window model at backcasting initial values is reassuring in the sense that there is sufficient information regarding the time variation in the data. However, across the full sample, we find that a longer estimation window length of 200 days improves the in-sample performance of a 100-days rolling window model.

[5]We make use of the Fortran implementation within the Model Selection Toolbox, available at https://github.com/aboisbunon/mst (accessed on April 30, 2014). Following the suggestion of Boisbunon et al. (2014), the penalty parameter λ is chosen to minimize the unbiased loss estimator $\delta_0(\beta_i) := \sum_{t=1}^{T-1}(y_{t+1,i} - y_t^\top \beta_i)^2 + (2\sum_{p=1}^{n} \|\beta_{i,p}\|_0 - T - 1) \times (T-1)^{-1} \sum_{t=1}^{T-1}(y_{t+1,i} - y_t^\top \widehat{\beta}_i)^2$.



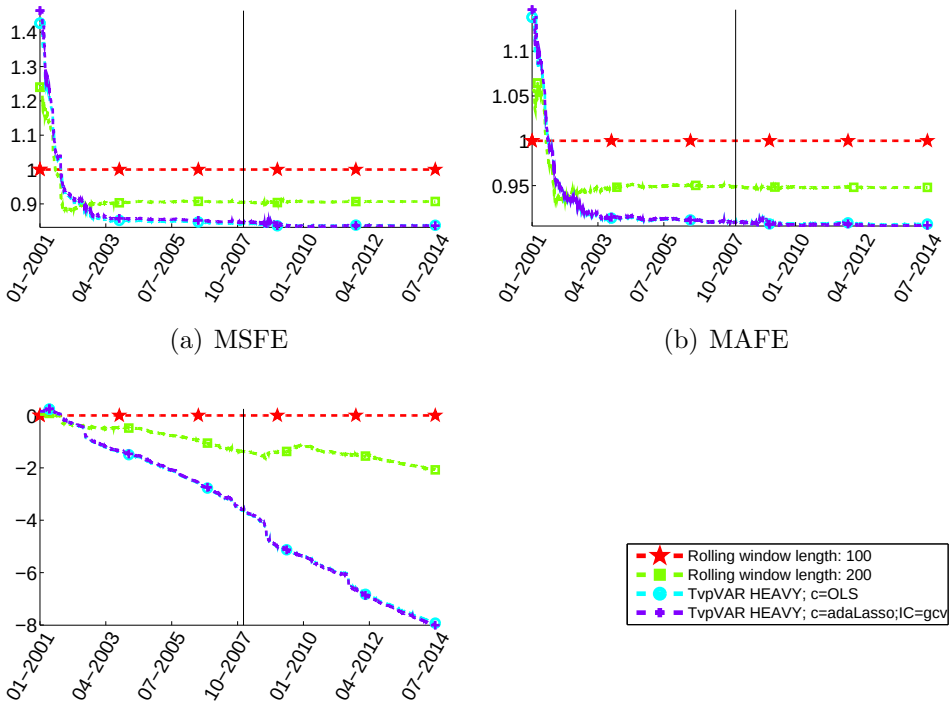
(a) MSFE (b) MAFE

(c) Negative QLikl (d) Legend

Figure 5.1: Relative forecasting performance comparison across specific assets

Notes: The solid lines depict the in-sample fit, while the dotted lines depict the out-of-sample fit. Panels (a) and (b) report $\left(\sum_{s=1}^{T} l_{s,\text{Alternative}}\right) / \left(\sum_{s=1}^{T} l_{s,\text{Benchmark}}\right)$ with $l_s \in \{\text{MSFE}_s, \text{MAFE}_s\}$, while panel (c) reports $(1/T)\sum_{s=1}^{T}\left(l_{s,\text{Alternative}} - l_{s,\text{Benchmark}}\right)$ with $l_s = -QLikl_s$.

One might argue that the complex score-driven updating dynamics might over-parametrize the fit to the data, and we therefore additionally conduct an out-of-sample forecasting comparison. The out of sample forecast comparison begins with January 2, 2008. Such a choice serves two purposes. First, the critical litmus test of a systemic risk model lies in its predictive ability during the recent financial crisis. Second, we aim to remain conservative regarding the performance of our newly



(a) MSFE (b) MAFE

(c) Negative QLikl (units in thousands) (d) Legend

Figure 5.2: Relative forecasting performance comparison across time

Notes: The start of the out-of-sample evaluation period on January 1, 2008 is depicted by a horizontal line. Panels (a) and (b) report $\left(\sum_{s=1}^{t} l_{s,\text{Alternative}}\right) / \left(\sum_{s=1}^{t} l_{s,\text{Benchmark}}\right)$ with $l_s \in \{\text{MSFE}_s, \text{MAFE}_s\}$, while panel (c) reports $\sum_{s=1}^{t} \left(l_{s,\text{Alternative}} - l_{s,\text{Benchmark}}\right)$ with $l_s = -QLikl_s$.

introduced filtering model. By opting for a relatively short in-sample period, we potentially harm our parametric filtering model due to potential model instabilities and the increased parameter uncertainty.

One clear stylized fact that emerges is that the filtering models outperform all rolling window competitors, even in an out-of-sample context as documented by the pairwise forecasting comparisons in Table 5.1. Furthermore, we explore the effects of

a LASSO regularization of the steady state parameters. While the in-sample fit is slightly impaired by the penalized estimation, we do find rather modest benefits in the out of sample context. Summing up, we find that estimating the exponential decay of temporally weighted innovations entails sizable forecasting gains when compared to a rolling-window procedure.

Pairwise forecast comparisons in Table 5.1 reveal that all of the comparisons turn out statistically significant, which is perhaps not surprising due to the high-dimensional nature of our setting. It is therefore important to note that in terms of relative mean squared forecast errors (MSFE) and relative mean absolute forecast errors (MAFE), the score based filtering procedure outperforms the rolling window benchmark by ten percentage points.[6]

[6]The loss functions at time t for model i are defined as follows

$$\begin{aligned}
\text{MSFE}_{t,i} &= \left(y_{t,i} - y_{t-1}^{\top} f_{t,i}\right)^2 \\
\text{MAFE}_{t,i} &= \left|y_{t,i} - y_{t-1}^{\top} f_{t,i}\right| \\
QLikl_{t,i} &= \frac{1}{2}\left[\log\left(\sigma_{t,i}^2\right) + \left(\frac{y_{t,i} - y_{t-1}^{\top} f_{t,i}}{\sigma_{t,i}}\right)^2\right].
\end{aligned}$$

Table 5.1: Pairwise Giacomini-White predictive comparisons

Notes: Benchmark models are depicted in columns, while the alternative forecasting competitors are reported in rows. The non-parenthesized entries report differences in losses between benchmark and competitor models. The parenthesized p-values correspond to that of a formal forecasting comparison test of Giacomini and White (2006).

	Rolling window length: 100			Rolling window length: 200			TvpVAR HEAVY; c=OLS		
	negQLIKL	MSFE	MAFE	negQLIKL	MSFE	MAFE	negQLIKL	MSFE	MAFE
Rolling window length: 200	-695.270^{*}	-11890.711^{*}	-1907.933^{*}						
(P-value)	(0.001)	(0.000)	(0.000)						
TvpVAR HEAVY; c=OLS	-4311.715^{*}	-21546.419^{*}	-3440.667^{*}	-3616.445^{*}	-9655.708^{*}	-1532.734^{*}			
(P-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			
TvpVAR HEAVY; c=adaLASSO	-4410.571^{*}	-22170.034^{*}	-3538.016^{*}	-3715.301^{*}	-10279.323^{*}	-1630.083^{*}	-98.856^{*}	-623.615^{*}	-97.349^{*}
(P-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.029)	(0.004)

5.3.3 Spillover implications of alternative VAR models

The spillover plots and tables of Diebold and Yılmaz (2014) visualize the variance decompositions d_{kj}^H in (5.1). At every point in time, we compute the instantaneous variance decompositions from the time-varying location and scale parameters of the instantaneous VAR model at time t.[7] The purpose of the spillover decomposition framework is to quantify spillover risks in a timely fashion. We therefore measure spillovers based on daily data, refraining from time-aggregation to lower frequencies.[8] The horizon H for the variance decomposition in (5.1) is chosen to reflect the immediate timespan of the Basel III framework for short term risk reporting over the next ten business days ($H = 10$).

Table 5.2 reports the parameter estimates for alternative filtering models, highlighting the high degree of persistence in the time-varying parameters. One potential issue that may arise in practice is that for the impulse response functions to be well behaved for any arbitrary (possibly infinite) horizon H, the *instantaneous* VARs need to be stationary. However, we find mean reversion in the estimated dynamic autoregressive matrix. Much of the variation is explained by the own autoregressive lag,

[7] Reporting instantaneous variance decompositions provides an alternative to time consuming simulation based techniques under a high persistence assumption (see also Giraitis et al., 2014). However, we recognize that the omittance of time-varying parameter mean-reversion may overstate changes in spillover measures.

[8] The financial connectedness homepage http://financialconnectedness.org gives a comprehensive overview of various identification-related robustness checks.

Table 5.2: Hyperparameter estimates of the return spillover model

Notes: Estimates of the time-varying VAR model. The long run autoregressive matrix of model with "c=OLS" is estimated by OLS, while the long run autoregressive matrix of model with "c=adaLASSO" is estimated by adaptive LASSO. Standard errors are in parentheses.

	TvpVAR HEAVY c=OLS	TvpVAR HEAVY c=adaLASSO
α	0.0019	0.0019
	(0.0028)	(0.0074)
β	0.8203	0.8419
	(0.2059)	(0.2677)
γ^{σ}	1.3651	1.3757
	(0.0559)	(0.0582)
α_1^{σ}	0.0195	0.0194
	(0.0037)	(0.0044)
α_2^{σ}	0.0994	0.0996
	(0.0204)	(0.0212)
β^{σ}	0.9892	0.9900
	(0.0043)	(0.0042)
QLikl	$-23,256$	$-23,277$

facilitating diagonal dominance of the autoregressive matrix, such that the VARs indeed turn out to be stationary at every single point in time, see also Figure 5.3.[9] Moreover, the economic interest in our applications lies in a short risk horizon H, such that the asymptotic explosiveness potential of a non-stationary VAR is of little practical importance over a short horizon.

The unconditional descriptions in Tables 5.4 and 5.3 indicate that JP Morgan Chase and Alcoa are identified as the most central assets in terms of their spillover centrality. However, time variation in the spillover centrality of both of these assets is of a different nature: while Alcoa is an industry leader at the source of the supply

[9]If in further applications of our model the instantaneous VAR turns out to be non-stationary, we advise the practitioner to adopt a regularized LASSO-VAR as the off-diagonal zero coefficients effectively enforce lower eigenvalues of the autoregressive matrix. Furthermore, explicit parametrizations to enforce stationarity are described by Giraitis et al. (2014).

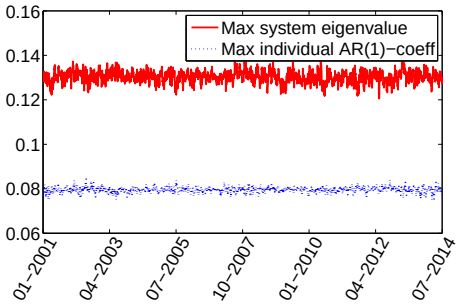


Figure 5.3: System and equation-by-equation instantaneous absolute eigenvalues of the autoregressive matrix in a TvpVAR model

Table 5.3: Unconditional return spillover centralities, over a horizon of 10 days

Notes: 'To' column corresponds to out-degree and 'From' corresponds to in-degree of the variance decomposition network.

	Spillover Centrality	To (most influential)	From (most vulnerable)	Net
AA	17.6	209.0	56.3	152.6
AXP	14.5	172.1	69.9	102.1
BA	5.8	69.3	77.8	−8.5
CAT	8.3	97.9	77.7	20.2
GE	8.2	96.9	81.1	15.8
HD	6.2	74.0	76.9	−2.9
HON	7.7	91.8	80.2	11.7
IBM	2.9	34.1	86.5	−52.3
JPM	17.1	203.2	58.6	144.6
KO	1.2	14.3	89.4	−75.1
MCD	1.5	17.8	82.0	−64.3
PFE	2.7	32.5	82.7	−50.1
PG	1.0	12.2	91.2	−79.0
WMT	1.7	19.7	88.3	−68.6
XOM	3.5	41.6	87.8	−46.2

chain, spillover impact of JP Morgan is focused around the financial crisis. We next explore further conditional spillover properties of the different models across different assets.

Table 5.4: Unconditional return spillover table, over a horizon of 10 days

	AA	*AXP*	*BA*	*CAT*	*GE*	*HD*	*HON*	*IBM*	*JPM*	*KO*	*MCD*	*PFE*	*PG*	*WMT*	*XOM*	From
AA	43.7	9.5	4.3	8.1	5.6	3.4	5.7	1.9	10.8	0.6	0.9	1.6	0.5	0.7	2.8	56.3
AXP	12.6	30.1	4.0	6.4	7.2	4.6	5.3	2.1	20.6	0.6	1.0	1.8	0.6	1.0	2.1	69.9
BA	16.0	11.3	22.2	7.4	6.5	4.5	9.4	2.0	11.8	0.8	1.3	2.0	0.7	1.2	2.9	77.8
CAT	19.7	11.7	4.8	22.3	6.6	4.4	7.5	2.2	13.2	0.7	0.9	1.7	0.5	1.0	2.8	77.7
GE	14.5	13.9	4.5	7.0	18.9	4.8	7.0	2.7	18.9	0.7	1.0	2.1	0.6	1.1	2.2	81.1
HD	12.2	12.7	4.4	6.6	6.8	23.1	6.2	2.6	16.0	0.9	1.4	1.8	0.7	2.5	2.1	76.9
HON	16.1	11.3	7.1	8.7	7.7	4.8	19.8	2.9	13.8	0.8	1.0	1.7	0.6	1.1	2.6	80.2
IBM	15.0	12.4	4.3	7.0	8.1	5.5	7.9	13.5	17.4	0.9	1.2	1.9	0.7	1.4	2.7	86.5
JPM	10.7	15.2	3.1	5.4	7.2	4.3	4.8	2.2	41.4	0.5	0.7	1.5	0.5	0.8	1.7	58.6
KO	15.5	12.3	5.6	7.1	6.5	6.1	6.9	2.7	12.9	10.6	1.8	3.6	2.2	1.9	4.5	89.4
MCD	14.4	12.1	5.8	6.3	6.4	6.3	5.9	2.5	12.4	1.2	18.0	2.5	1.1	1.9	3.2	82.0
PFE	14.9	12.7	5.0	6.4	7.4	4.5	5.6	2.2	14.2	1.3	1.4	17.3	1.3	1.7	4.1	82.7
PG	14.5	13.1	5.5	6.4	7.2	5.5	6.2	2.6	14.1	2.7	1.9	4.2	8.8	2.2	5.0	91.2
WMT	11.6	12.2	5.3	6.6	7.3	11.0	6.7	3.1	14.4	1.2	1.9	3.0	1.2	11.7	3.0	88.3
XOM	21.2	11.8	5.6	8.5	6.3	4.2	6.7	2.6	12.7	1.3	1.4	3.2	1.2	1.3	12.2	87.8
To	209.0	172.1	69.3	97.9	96.9	74.0	91.8	34.1	203.2	14.3	17.8	32.5	12.2	19.7	41.6	1186.5
CiO	252.6	202.1	91.5	120.2	115.8	97.1	111.7	47.7	244.6	24.9	35.7	49.9	21.0	31.4	53.8	1500.0

Notes: Abbreviations: To $\triangleq$ Contribution To Others; From $\triangleq$ Contribution from Others; CiO $\triangleq$ Contribution Including Own.

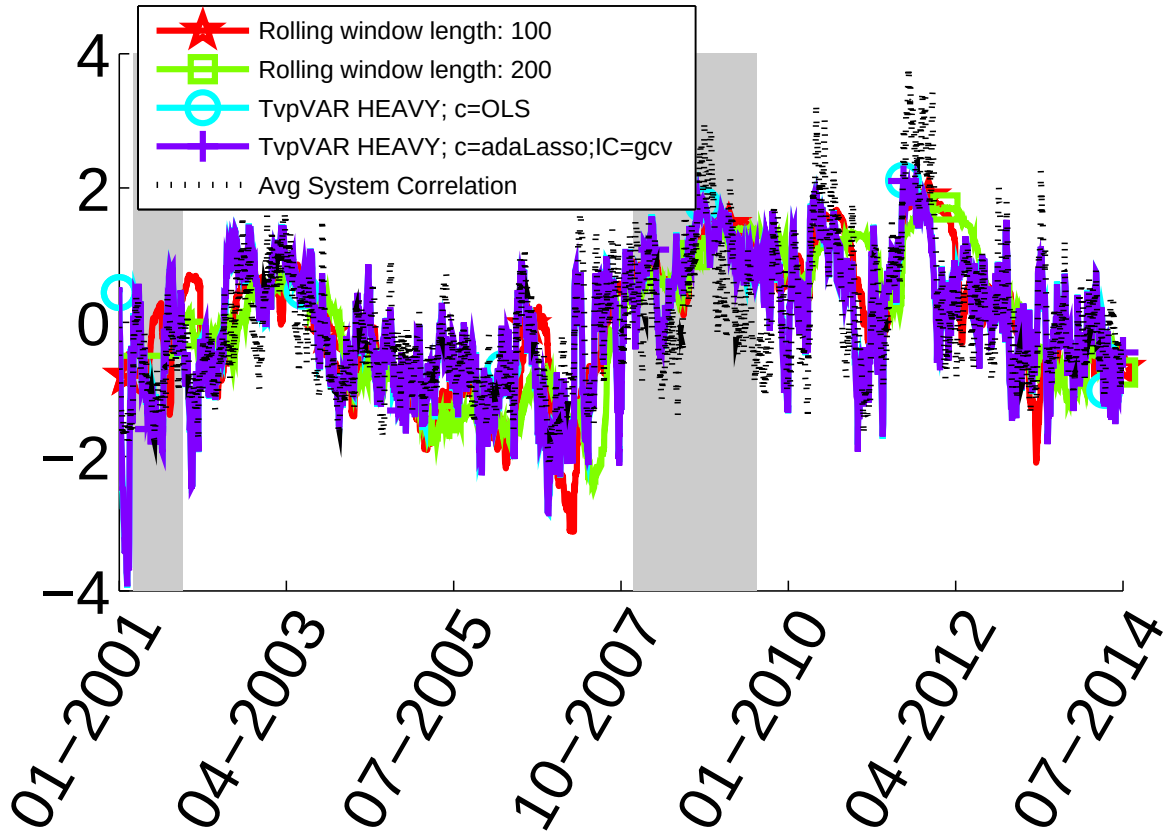


Figure 5.4: Transformed spillover index time series

Notes: In order to make the systemic risk indices comparable, we transformed the series to have mean zero and standard deviation one. The colored lines correspond to standardnormalized spillover indices of competing models, see also panel (d) of Figure 5.6. In particular, the red line corresponds to 100-day rolling window estimation, the green line corresponds to a 200-day rolling window estimation, the light blue line corresponds to the TvpVAR model, and the purple line corresponds to the TvpVAR model with LASSO-regularized estimation. The black dotted line depicts the dynamic equicorrelation among equity returns, see also Chapter 3.

Our approximating VAR model captures time-variation in the return spillover index. Interestingly, relative time-variation in the spillover indices coincides with the relative time-variation in correlation levels of a simple equicorrelation model. The Pearson correlation coefficient between the spillover index of a TvpVAR model and the dynamic equicorrelation is 0.89. This is displayed in Figure 5.4. Our observation that spillover measures and correlations move together addresses a common methodological concern within systemic risk measurement: by not modeling the mean equation and merely using covariance matrices to make statements regarding financial stability, sequential domino effects might mistakenly be attributed to coinciding contemporaneous comovement patterns. Even though sequential externalities and the contemporaneous comovement may have distinct policy implications, it is reassuring to observe the close empirical similarity of both channels.

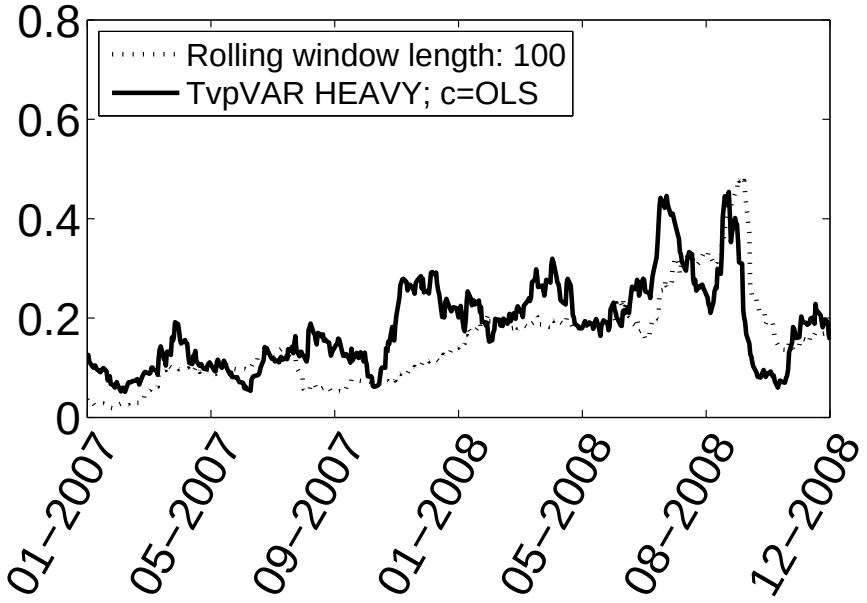
To gain insight into the dynamic nature of the individual spillover sources, we study the individual contributions to spillover measures. Comparing the implications alternative modeling frameworks for all individual assets in Figure 5.9 of Appendix 5.B, we find the filtering models adapt more quickly to the changing financial conditions, while the rolling window models extract more persistent spillover indexes. Such a feature enables the filtering models to identify the changes in spillovers in a more timely manner. If the policy maker desires a more persistent spillover source decomposition that would remain robust to outliers while retaining a reasonable forecasting performance, we would suggest one of the two adjustments to our methodology: di-

rectly smoothing out the model-implied spillover measurements by a moving average recursion; or replacing the innovations term in (5.4) by an average of lagged scores, say $(1/10) \times \sum_{i=1}^{10} s_{t-i+1}$, see also Patton (2006b).

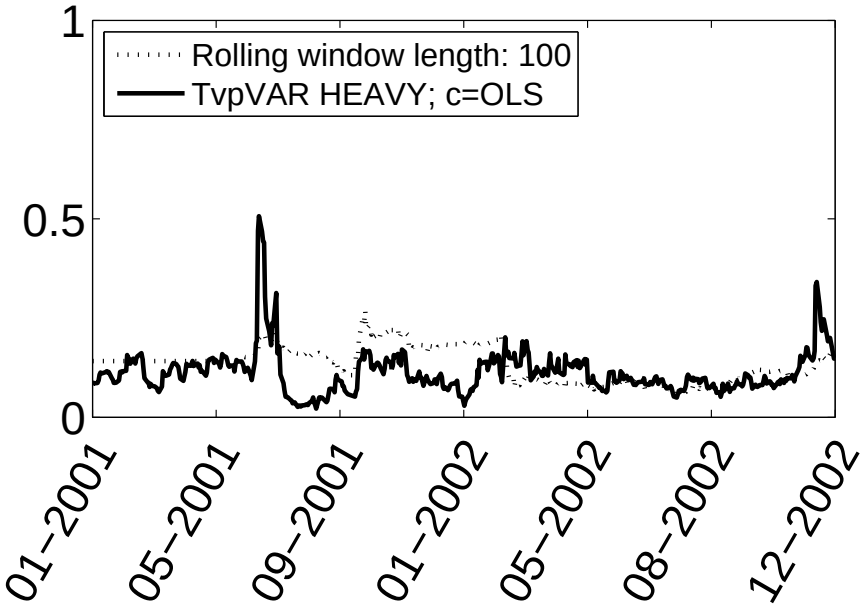
An obvious point of interest is to study the implications of both models during the onset of the financial crisis. We now discuss the spillover features of different models on the basis of two prominent examples. As a first example, we notice how all models identify JP Morgan Chase as the central asset in late 2008 because JPM is the only financial stock in our asset universe. The individual spillover centralities are plotted in Figure 5.5. We observe that the filtering model identifies JP Morgan as the source of spillovers more early than the rolling window approach. Taking a closer look, we see that the time-varying parameter specification detects two peaks in the spillover centrality of JP Morgan Chase, following the fall of Lehman Brothers on Sept 15, 2008 and following the asset purchases of the Troubled Assets Relief Program (TARP) on October 28, 2008. The rolling window alternative on the other hand maintains high levels of spillover centrality of JP Morgan Chase for a long time, even after the market turmoil has come to an end.

As a second tale for spillovers in equity market returns, we study the inferred high spillover centrality of Honeywell in 2001. In July 3, 2001 the European Commission stopped the merger of General Electric and Honeywell due to concerns about market competition. This may have raised some concerns regarding the regulatory environment, causing some short-term movements in the market. However, it is doubtful

whether Honeywell as an individual company ignites major market movements. In retrospect, it turns out that this episode may be classified a "false alarm" in our systemic risk measurement framework. Such an episode can pinpoint the major differences of the alternative modeling frameworks. While the spillover centrality of Honeywell in the score driven framework is indeed short-lived, the rolling window alternative downgrades the systemic relevance of Honeywell only once the single market moving observation exits the rolling window sample.



(a) JP Morgan Chase



(b) Honeywell International

Figure 5.5: Series of spillover centralities for key assets

5.4 Volatility spillovers

In our context of global equity markets, high-frequency volatility measures are publicly available from the Oxford Man Institute (Heber et al., 2009). We study daily realized variance measures over the period from July 8, 2002 to June 9, 2015, with a pseudo out-of-sample predictive analysis starting from January 2, 2008. We choose the 5-minute realized variance as the relevant volatility proxy due to its empirical success across a wide range of applications (Liu et al., 2015).[10] To ensure approximate normality in our system variables, we take logs of realized variances, denoted by $y_t = \log(RV_t)$. We see from the descriptive statistics in Table 5.10 of Appendix 5.B that the kurtosis of log variances is indeed much lower than in the equity returns applications.

The asset universe of international stock indices is listed in Table 5.10 of Appendix 5.B. In our international dataset, we find quite a few missing observations.[11] There are several economic sources that may give rise to missing observations. First, banking holidays differ across different countries. Second, missing observations may arise especially due to market frictions in the context of financial crises. Third, if we observe a zero realized variance measure that may come about due to low liquidity, we can avoid taking logs of zeros by treating the observation as missing – as opposed to hardcoding the realized variance to a small number.

[10] Note that in the context of international markets, realized proxies for covariance matrices are in general not available due to largely non-overlapping trading hours.

[11] The missing observations are depicted in Figure 5.10 of Appendix 5.B.

The advantage of our structural time series formulation is that one is able to take informational advantage from unbalanced panels (increased information in equity returns due to a longer data span). Handling the frequently occurring missing observations is straightforward as the implied score for coefficients with missing information can be set to zero (Creal et al., 2014; Durbin and Koopman, 2012). We adjust score based updating to account for missing observations as described in Appendix 5.C. For the rolling window benchmark, missing observations are less straightforward to handle. We therefore deal with missing observations by linearly interpolating the missing observations for the rolling window model, which again acts as a conservative mechanism to aid the benchmark with artificial additional data from the future.

5.4.1 Multivariate scale dynamics

Spillover dynamics at a high frequency can appear as contemporaneous correlation patterns when the data is sampled at a lower frequency. For this reason care needs to be taken in order to account for variation in second moments. We do not find strong evidence of time-varying volatility of volatility and therefore persevere with modeling dynamic correlations. The purpose of allowing for the residuals correlation is to absorb dependence on the contemporaneous market factor. The equicorrelation matrix structure from Chapter 3 captures a simple factor structure. In short, the variance model in our spillover system (5.2)– (5.5) is specified by the usual VAR(1)-

system for location dynamics and the following model for variance matrix

$$\Sigma_t = D_t R_t D_t$$

$$D_t = \text{diag}\left(\{\sigma_i\}_{i=1,\dots,n}\right)$$

$$R_t = \begin{pmatrix} 1 & \rho_t & \dots & \rho_t \\ \rho_t & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_t \\ \rho_t & \dots & \rho_t & 1 \end{pmatrix},$$

where ρ_t is driven by GAS(1,1) dynamics, see also Chapter 3 for further exposition.

The modeling choice for the dynamic variance matrix can as well be interpreted as a particular spillover structure in which all assets mutually affect each other in equal proportions. We opt for this specification as it helps capture the commonality between assets, while not distorting the spillover interpretations of the model. Alternatively, contemporaneous factors can be accounted for by incorporating these into the measurement equation or by considering a model for the residuals of a pre-specified factor model (see e.g. Giglio et al., 2015).

5.4.2 Statistical evaluation of alternative VAR models

We estimate all filtering models via maximum (quasi-)likelihood. Due to the lack of realized (co-)volatility of volatility (quarticity) measures, we are required to estimate the volatility and correlation as specified in the previous subsection. Due to the inherent non-linearities in the model, we do this in a joint single-step estimation procedure. The TVP-VAR model is estimated via Gaussian maximum likelihood, while for adaptive LASSO estimation, we maximize the penalized criterion function and then report the Gaussian likelihood for comparison purposes.

The parameter estimates are reported in Table 5.5. We find a somewhat higher persistence in the dynamics of the autoregressive matrix when compared to the persistence in correlations. The implications for the spillover measurement indexes are discussed further below.

Table 5.5: Hyperparameter estimates of the volatility spillover model

Notes: Estimates of the time-varying VAR model. The long run autoregressive matrix of model with "c=OLS" is estimated by OLS, while the long run autoregressive matrix of model with "c=adaLASSO" is estimated by adaptive LASSO. The parameters ω^ρ, α^ρ, β^ρ are the equicorrelation GAS intercept, score and autoregressive parameters, respectively. Standard errors are in parentheses.

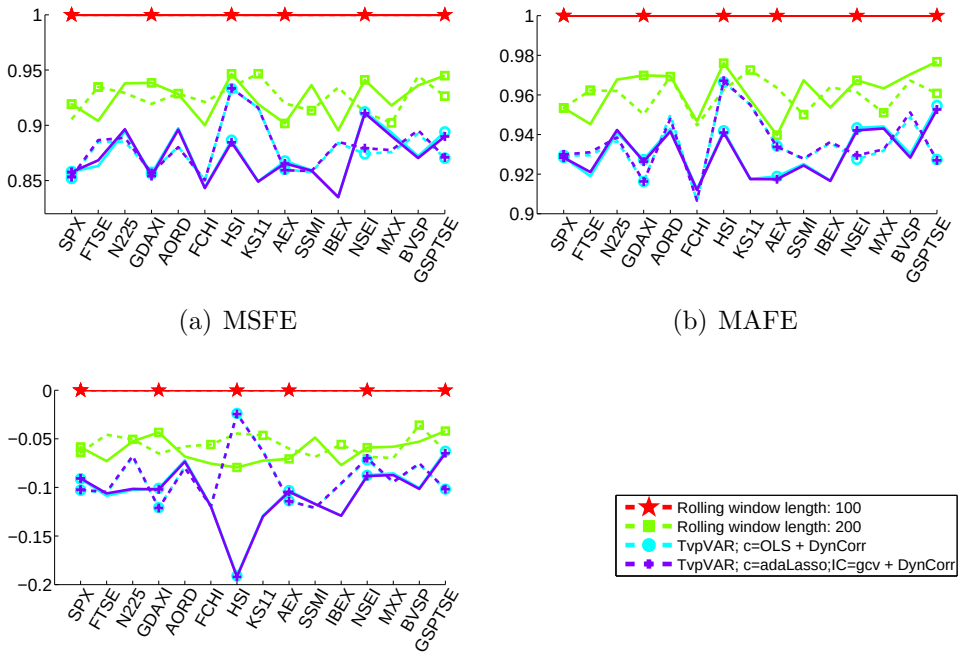
	TvpVAR c=OLS	TvpVAR c=adaLASSO
α	0.0072	0.0074
	(0.0009)	(0.0008)
β	0.9976	0.9978
	(0.0009)	(0.0008)
σ_{SPX}	0.5811	0.5796
	(0.0170)	(0.0247)
σ_{FTSE}	0.5225	0.5217
	(0.0194)	(0.0238)
σ_{N225}	0.5384	0.5378
	(0.0145)	(0.0222)
σ_{GDAXI}	0.5338	0.5318
	(0.0164)	(0.0221)
σ_{AORD}	0.6759	0.6733
	(0.0219)	(0.0315)
σ_{FCHI}	0.5024	0.5008
	(0.0176)	(0.0222)
σ_{HSI}	0.5749	0.5729
	(0.0196)	(0.0266)
σ_{KS11}	0.5527	0.5506
	(0.0177)	(0.0256)
σ_{AEX}	0.5109	0.5090
	(0.0182)	(0.0239)
σ_{SSMI}	0.4384	0.4372
	(0.0162)	(0.0213)
σ_{IBEX}	0.5263	0.5249
	(0.0186)	(0.0241)
σ_{NSEI}	0.6935	0.6929
	(0.0278)	(0.0364)
σ_{MXX}	0.6936	0.6916
	(0.0229)	(0.0294)
σ_{BVSP}	0.5545	0.5538
	(0.0167)	(0.0232)
σ_{GSPTSE}	0.5667	0.5652
	(0.0161)	(0.0234)
ω^ρ	0.0145	0.0578
	(0.1261)	(0.3270)
α^ρ	0.0272	0.0231
	(0.0128)	(0.0317)
β^ρ	0.9409	0.9618
	(0.0633)	(0.1215)
QLikl	2,886	2,881

The statistical evaluation of alternative volatility spillover models bears some similarity with the patterns in return spillovers. We therefore briefly sketch the patterns to focus on the economic interpretations of the competing models further below. The time frame of our study spans the period between July 8, 2002 and June 9, 2015 with an out of sample forecast comparison beginning with January 2, 2008.

We also observe how the filtering model delivers superior short-term forecasts of volatility spillovers both in-sample and out-of-sample, across time and across different assets. Table 5.6 reports a consistent forecasting ranking of alternative models according to all loss functions used (MAFE, MSFE, negQLIKL). The 100-day rolling window benchmark is dominated by the 200-day rolling window estimation. Furthermore, we find that the filtering models outperform the rolling window schemes. Differences between filtering models are relatively small: the time-varying parameter filtering model without regularization performs similarly as the time-varying parameter filtering model that uses LASSO estimation of the intercept.[12] Figure 5.6 disaggregates the advantage in the forecasting performance into the cross section of assets. We observe that the filtering model improves upon the benchmarks for almost all assets – only for Hong-Kong and Korean stock exchanges the out-of-sample comparison comes close. Figure 5.7 displays the forecasting comparison across time. In the in-sample evaluation, the rolling window model only outperforms near the initial values. Recall that in order to avoid numerically optimizing over n^2 intercept parameters for the filtering

[12] A Giacomini-White statistical forecasting test suggest that the LASSO estimation slightly outperforms the standard filtering benchmark in the current out of sample context, yet the interpretation of statistical significance is problematic due to high power of tests in large datasets.

model, we have estimated the intercept in the dynamic VAR updating recursion (5.3) via a targeting procedure. The unconditional autoregressive matrix is also used as the initial value for the time-varying VAR parameters. As the filtering model obtains the steady dynamics over the entire in-sample period, data from around the initial period is more informative for explaining the dynamics within the asset universe of our study. Such a pattern is reassuring in the sense of indicating time-variation in the behavior of the predictive system. We see that the advantage of the filtering model is valid throughout the entire hold-out sample.



(a) MSFE

(b) MAFE

(c) Negative QLikl

(d) Legend

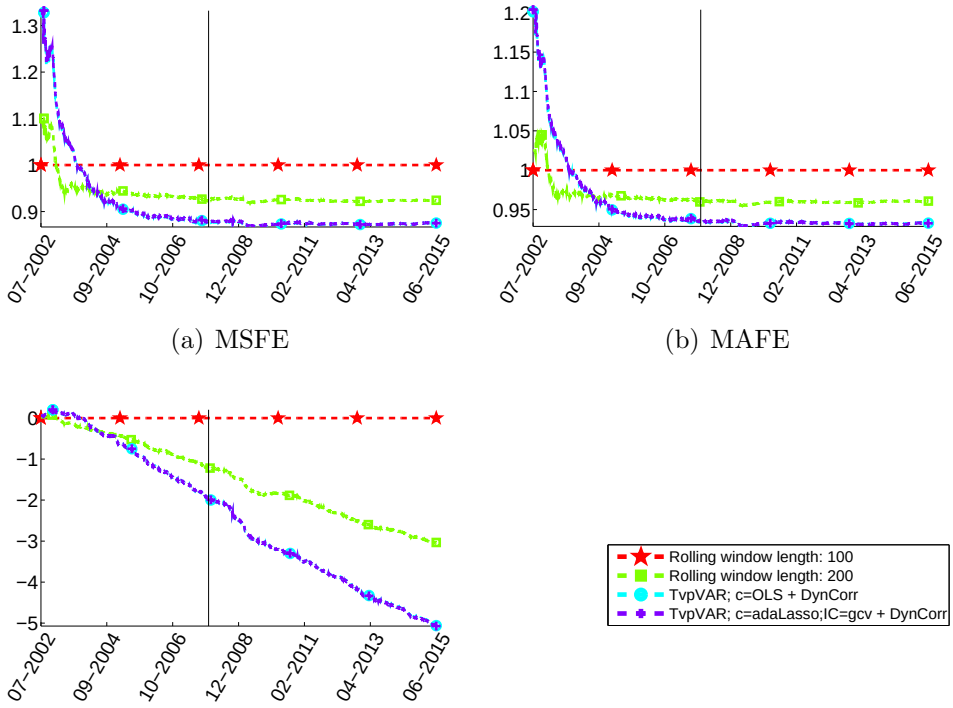
Figure 5.6: Relative forecasting performance comparison across specific assets

Notes: The solid lines depict the in-sample fit, while the dotted lines depict the out-of-sample fit. Panels (a) and (b) report $\left(\sum_{s=1}^{T} l_{s,\text{Alternative}}\right) / \left(\sum_{s=1}^{T} l_{s,\text{Benchmark}}\right)$ with $l_s \in \{\text{MSFE}_s, \text{MAFE}_s\}$, while panel (c) reports $(1/T)\sum_{s=1}^{T}\left(l_{s,\text{Alternative}} - l_{s,\text{Benchmark}}\right)$ with $l_s = -QLikl_s$.

Table 5.6: Pairwise Giacomini-White predictive comparisons

Notes: Benchmark models are depicted in columns, while the alternative forecasting competitors are reported in rows. The non-parenthesized entries report differences in losses between benchmark and competitor models. The parenthesized p-values correspond to that of a formal forecasting comparison test of Giacomini and White (2006).

	Rolling window length: 100			Rolling window length: 200			TvpVAR; c=OLS		
	negQLIKL	MSFE	MAFE	negQLIKL	MSFE	MAFE	negQLIKL	MSFE	MAFE
Rolling window length: 200	-1817.703^{*}	-739.389^{*}	-485.227^{*}						
(P-Value)	(0.000)	(0.000)	(0.000)						
TvpVAR; c=OLS	-3116.891^{*}	-1239.205^{*}	-867.300^{*}	-1299.188^{*}	-499.816^{*}	-382.073^{*}			
(P-Value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			
TvpVAR; c=adaLASSO	-3131.583^{*}	-1249.137^{*}	-875.038^{*}	-1313.880^{*}	-509.747^{*}	-389.811^{*}	-14.692^{*}	-9.931^{*}	-7.737^{*}
(P-Value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.100)	(0.051)	(0.037)



(a) MSFE

(b) MAFE

(c) Negative QLikl (units in thousands)

(d) Legend

Figure 5.7: Relative forecasting performance comparison across time

Notes: The start of the out-of-sample evaluation period on January 1, 2008 is depicted by a horizontal line. Panels (a) and (b) report $\left(\sum_{s=1}^{t} l_{s,\text{Alternative}}\right) / \left(\sum_{s=1}^{t} l_{s,\text{Benchmark}}\right)$ with $l_s \in \{\text{MSFE}_s, \text{MAFE}_s\}$, while panel (c) reports $\sum_{s=1}^{t} \left(l_{s,\text{Alternative}} - l_{s,\text{Benchmark}}\right)$ with $l_s = -QLikl_s$.

5.4.3 Spill-over implications of alternative VAR models

The unconditional spillover descriptions in Table 5.7 verify the centrality of the US stock market, followed by the Canadian market. Asian markets seem to be influenced

by European and North-American markets, while they have a less strong impact on other markets. Furthermore, we see that the more spillover central markets also have positive net contribution to the rest of the system. The less spillover central Asian markets are on the receiving end of the shocks.

Table 5.7: Unconditional volatility spillover centralities, over a horizon of 10 days

Notes: 'To' column corresponds to out-degree and 'From' corresponds to in-degree of the variance decomposition network.

System	Country	Spillover Centrality	To (most influential)	From (most vulnerable)	Net
SPX	US	16.0	171.2	72.2	99.0
GSPTSE	Canada	11.8	126.4	66.7	59.7
FTSE	FTSE	10.0	107.0	84.9	22.1
AEX	Netherlands	9.3	99.2	85.5	13.7
GDAXI	Germany	9.1	96.9	80.8	16.1
FCHI	France	8.5	90.9	86.8	4.1
MXX	Mexico	7.6	81.5	56.0	25.5
SSMI	Switzerland	6.7	71.7	88.0	−16.3
IBEX	Spain	6.5	70.1	79.0	−8.9
BVSP	Brazil	3.9	41.9	70.9	−28.9
AORD	Australia	2.8	29.5	74.6	−45.1
KS11	Korea	2.7	28.9	64.7	−35.8
N225	Japan	1.9	20.4	65.8	−45.4
NSEI	India	1.9	20.1	42.7	−22.6
HSI	China	1.4	14.9	52.0	−37.1

Table 5.8 decomposes the spillovers to a more granular level. We observe that the emerging markets have the most idiosyncratic behavior. This can be seen from the dominant role of own lagged shocks in explaining the variance of emerging markets' returns and from the relatively little influence on other markets. By contrast, open developed economies display a pattern of sharing a large portion of the variance with other markets. For instance, this is the case for SPX and FTSE markets.

Table 5.8: Unconditional volatility spillover table, over a horizon of 10 days

	SPX	FTSE	N225	GDAXI	AORD	FCHI	HSI	KS11	AEX	SSMI	IBEX	NSEI	MXX	BVSP	GSPTSE	From
SPX	27.8	8.8	1.4	7.9	1.8	7.2	0.9	2.6	7.9	5.7	5.3	0.6	6.6	3.7	11.6	72.2
FTSE	14.4	15.1	1.2	10.8	2.4	10.2	1.1	1.7	11.2	7.7	8.1	0.4	4.2	2.3	9.1	84.9
N225	10.8	6.7	34.2	5.1	1.6	4.4	1.5	3.8	5.9	5.3	2.7	1.8	6.1	3.3	6.9	65.8
GDAXI	13.2	10.9	1.2	19.2	1.0	11.7	0.6	1.9	12.8	8.1	8.6	0.2	2.0	2.4	6.5	80.8
AORD	11.3	6.2	1.2	3.4	25.4	4.5	1.0	0.7	4.0	4.0	5.2	1.1	13.8	2.7	15.4	74.6
FCHI	13.5	11.1	0.9	12.7	2.0	13.2	0.6	1.1	12.4	7.2	10.8	0.1	3.5	2.1	8.7	86.8
HSI	8.2	6.1	1.6	3.0	1.1	2.9	48.0	4.4	3.8	3.4	2.0	4.8	2.8	2.6	5.2	52.0
KS11	13.7	6.6	3.1	6.7	0.3	3.7	2.4	35.3	5.9	5.2	0.5	4.8	2.1	4.4	5.2	64.7
AEX	14.0	11.5	1.2	13.1	1.5	11.7	0.7	1.7	14.5	7.8	8.6	0.2	3.1	2.4	7.8	85.5
SSMI	14.3	11.5	1.9	11.5	2.2	9.7	1.0	2.2	11.2	12.0	7.2	0.5	4.1	2.5	8.1	88.0
IBEX	11.7	9.7	0.6	10.5	2.7	12.2	0.5	0.2	10.1	6.0	21.0	0.1	3.6	1.6	9.5	79.0
NSEI	6.0	2.9	2.0	1.3	1.2	1.0	2.5	4.3	2.0	1.9	0.1	57.3	7.2	3.8	6.4	42.7
MXX	11.7	3.8	1.4	1.7	5.0	2.7	0.7	0.7	2.4	2.3	2.6	2.1	44.0	4.6	14.3	56.0
BVSP	14.2	4.9	1.9	5.0	1.9	3.9	1.0	2.8	4.7	3.4	2.7	2.2	10.9	29.1	11.4	70.9
GSPTSE	14.1	6.1	0.7	4.2	4.9	5.3	0.6	0.8	4.8	3.5	5.6	1.0	11.3	3.6	33.3	66.7
To	171.2	107.0	20.4	96.9	29.5	90.9	14.9	28.9	99.2	71.7	70.1	20.1	81.5	41.9	126.4	1070.4
CiO	199.0	122.1	54.6	116.1	54.9	104.1	62.9	64.2	113.7	83.7	91.1	77.4	125.5	71.1	159.7	1500.0

Notes: To $\triangleq$ Contribution To Others; From $\triangleq$ Contribution from Others; CiO $\triangleq$ Contribution Including Own.

Conditional spillover properties are plotted in Figure 5.8. We observe also in this example how the GAS implied spillover measures adapt more quickly to changes in market conditions. In particular, the spillover measures of the filtering models have much less of a delay in returning to the pre-crisis levels.

When calculating spillover measures, we are required to make a choice regarding the contemporaneous covariance matrix entering the variance decomposition in (5.1). By having already accounted for time-varying correlation levels in our filtering model, we refrain from systematically overemphasizing periods of high volatilities and correlations. As the specific heterogeneous covariance structure does not enter the score driven updating recursion for the location parameters in (5.2) – (5.4), the impact of not having specified a heterogeneous correlation structure is to equally weigh each observation equation within the full system criterion function (see also Koopman et al., 2015). As an additional robustness check, we therefore estimate dynamic covariances as the final stage of a multistep estimation procedure by fitting Constant Conditional Correlation (CCC) and Dynamic Conditional Correlation (DCC) models to the standardized residuals. The alternative spillover indexes are plotted in Figure 5.11. We find that CCC and DCC models have relatively similar implications throughout, therefore it is reassuring to observe that the correlation model does not entirely drive the results. The dynamic equicorrelation implied spillover measures constitute an upper bound on the spillover measure. Hence, there is more room to diversification than a equicorrelation factor spillover structure would imply. Interestingly, we find

that the upper bound of the dynamic equicorrelation model is binding during recessions, while during macroeconomic expansions the diversification benefits are more substantial.

While there is rich time-variation in the dynamic VAR parameters, the spillover measures implied by a score-driven framework are much more stable than the rolling-window implied counterparts.[13] This could have to do with the high degree of persistence in realized volatility measures (Chiriac and Voev, 2011; Corsi, 2009). As the statistical performance of the filtering model is superior, it seems that a rolling window specification creates additional parameter uncertainty. The high degree of persistence in realized volatility implies that an observation more than 100 days ago still receives a relatively high weight ($\widehat{\beta}^{100} \approx 0.9978^{100} \approx 0.8$) within the GAS framework, while the rolling window approach cuts off all observations in further past. This results in superior performance of the 200-day rolling window when compared to a 100-day rolling window. However, within the GAS framework we are able to *estimate* the persistence of coefficient paths, such that the informational decay within the GAS model is much slower, effectively pooling more information into the filtered parameter paths.

[13] We report all spillover series for alternative models in Figure 5.12 of Appendix 5.B.

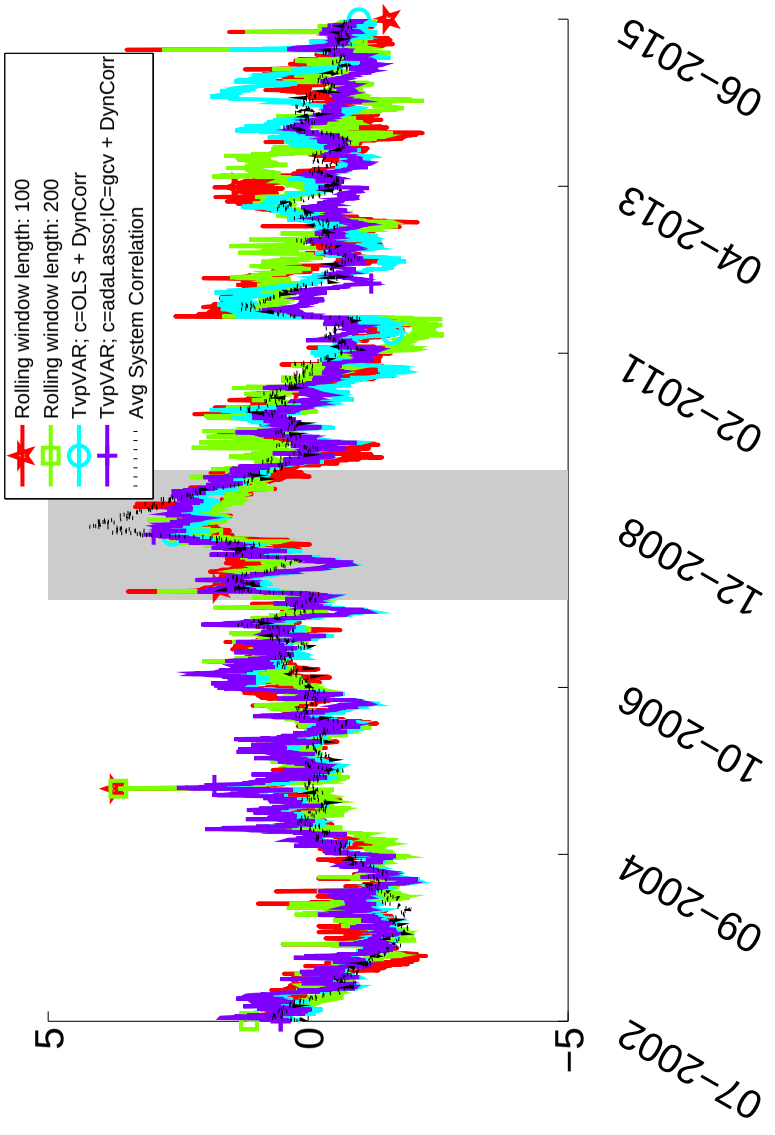


Figure 5.8: Transformed spillover index time series

Notes: The DCC model is used to describe the spillover measures. In order to make the systemic risk indices comparable, we transformed the series to have mean zero and standard deviation one. The colored lines correspond to standardnormalized spillover indices of competing models, see also panel (d) of Figure 5.7. In particular, the red line corresponds to 100-day rolling window estimation, the green line corresponds to a 200-day rolling window estimation, the light blue line corresponds to the TvpVAR model, and the purple line corresponds to the TvpVAR model with LASSO-regularized estimation. The black dotted line depicts the dynamic equicorrelation among equity returns, see also Chapter 3.

5.5 Conclusions

We monitor price spillover externalities in equity markets. We show how fully specified models for dynamic predictive systems can deliver superior forecasts and more precise spillover measures. In a statistical forecasting comparison, we find that most of the forecasting benefits stem from temporal weighting of the score-based innovations. While statistically significant, the gains of a LASSO-based regularization are modest in economic terms. From an economic perspective, the proposed approach delivers more timely indicators of financial ripple effects that could be driven by systemic risk.

There are further extensions that could be considered. For instance, an appealing alternative approach would be to consider time-varying factor augmented vector autoregressive (FAVAR) models to uncover the spillover dynamics. Such an approach would require one to work out the spillover interpretations of the FAVAR-based variance decomposition. In particular, if the factors are spanned by fundamental financial and macroeconomic quantities, then the increased magnitudes of spillovers in times of financial crises could be given the true meaning of contagion.

5.A Return Spillovers

Table 5.9: Descriptive statistics for daily US stock data

Symbol	Country	Kurtosis	Std	Min	Max
AA	Alcoa Inc.	10.760	2.693	−17.479	20.894
AXP	American Express Company	12.815	2.419	−19.370	18.754
BA	The Boeing Company	9.338	1.937	−19.407	14.360
CAT	Caterpillar Inc.	7.764	2.076	−15.728	13.693
GE	General Electric Company	11.865	2.011	−13.665	18.003
HD	The Home Depot	8.851	1.952	−15.178	13.145
HON	Honeywell International	11.111	2.001	−19.095	11.544
IBM	International Business Machines	9.700	1.581	−10.691	11.330
JPM	JP Morgan Chase	16.110	2.693	−23.235	22.385
KO	Coca-Cola	12.810	1.255	−10.611	12.990
MCD	McDonald's	9.583	1.457	−13.746	8.945
PFE	Pfizer	8.880	1.603	−11.803	9.701
PG	Procter & Gamble	9.757	1.183	−8.336	9.706
WMT	Wal-Mart Stores Inc.	7.949	1.365	−8.417	10.492
XOM	Exxon Mobil	14.952	1.575	−15.051	15.839

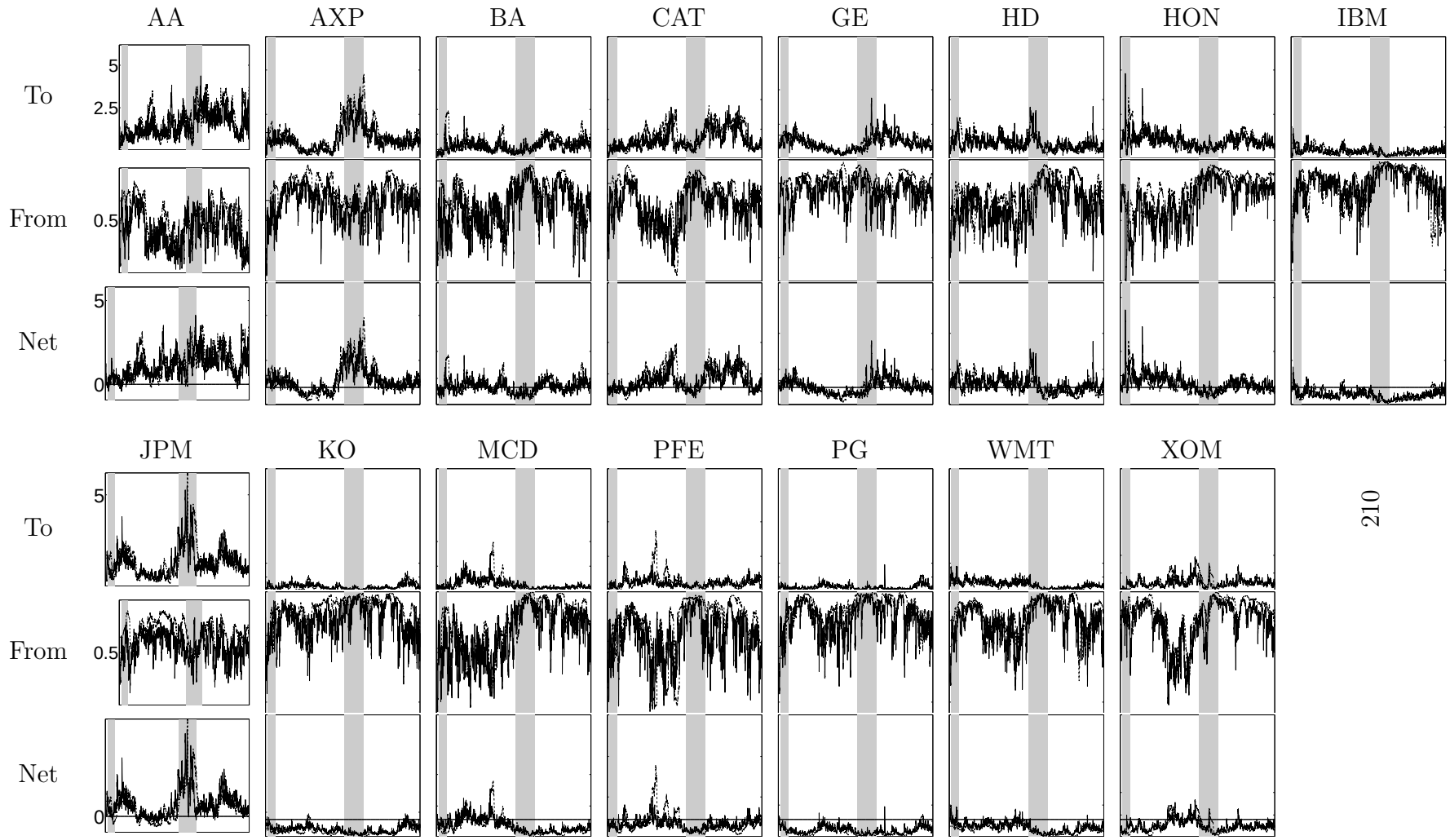


Figure 5.9: Individual return spillover series

Notes: The solid lines correspond to the TvpVAR model, while the dashed lines correspond to a rolling window of 100 days.

5.B Volatility Spillovers

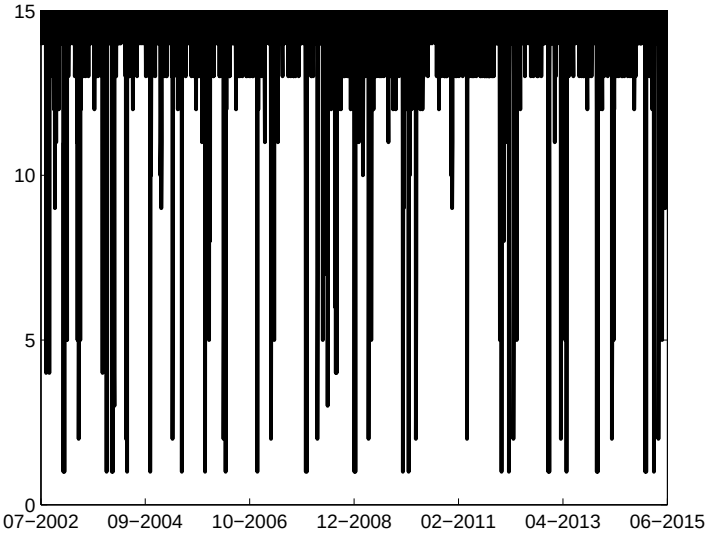


Figure 5.10: Number of available assets in international stock index data

Table 5.10: Descriptive statistics for international stock index data

Symbol	Country	Kurtosis	Std	Min	Max	# Missings
SPX	US	3.553	1.097	−3.557	4.915	141
FTSE	UK	3.217	1.049	−2.457	4.651	133
N225	Japan	3.991	0.857	−2.519	3.901	251
GDAXI	Germany	3.226	1.047	−2.893	4.169	108
AORD	Australia	3.465	0.975	−3.107	4.009	155
FCHI	France	3.096	1.002	−2.990	4.148	82
HSI	China	5.067	0.766	−2.372	4.388	451
KS11	Korea	3.507	0.940	−2.468	4.486	187
AEX	Netherlands	3.105	1.046	−3.382	4.005	82
SSMI	Switzerland	3.641	0.951	−1.975	4.390	146
IBEX	Spain	2.689	1.017	−3.069	3.820	109
NSEI	India	4.083	0.951	−3.074	5.291	193
MXX	Mexico	3.556	0.972	−2.653	4.521	136
BVSP	Brazil	4.879	0.773	−2.795	3.717	208
GSPTSE	Canada	3.982	1.052	−2.795	4.818	156

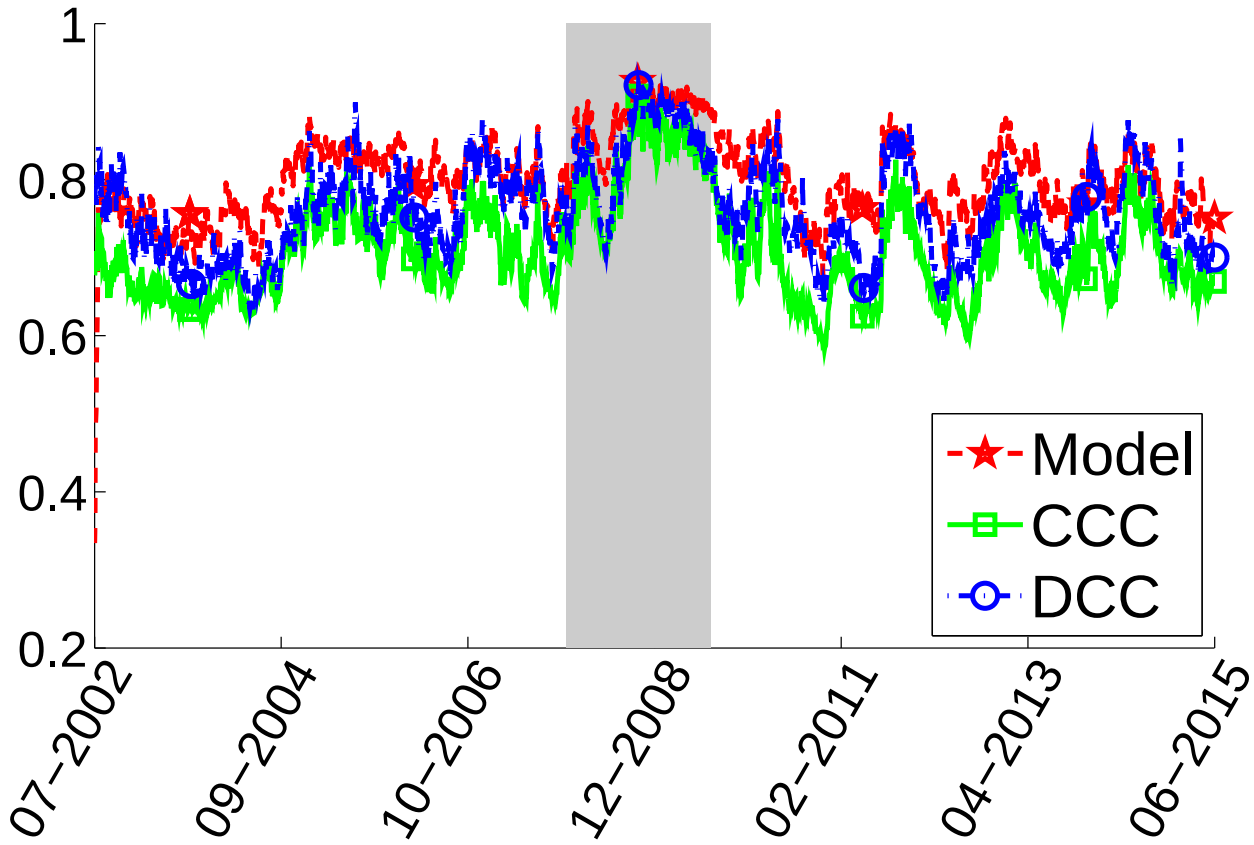


Figure 5.11: Spillover indices for alternative correlation model specifications

Notes: The underlying location dynamics are captured by a TvpVAR model with OLS intercept targeting. The alternative correlation models are: dynamic equicorrelation ("Model"), constant conditional correlation ("CCC") and dynamic conditional correlation ("DCC").

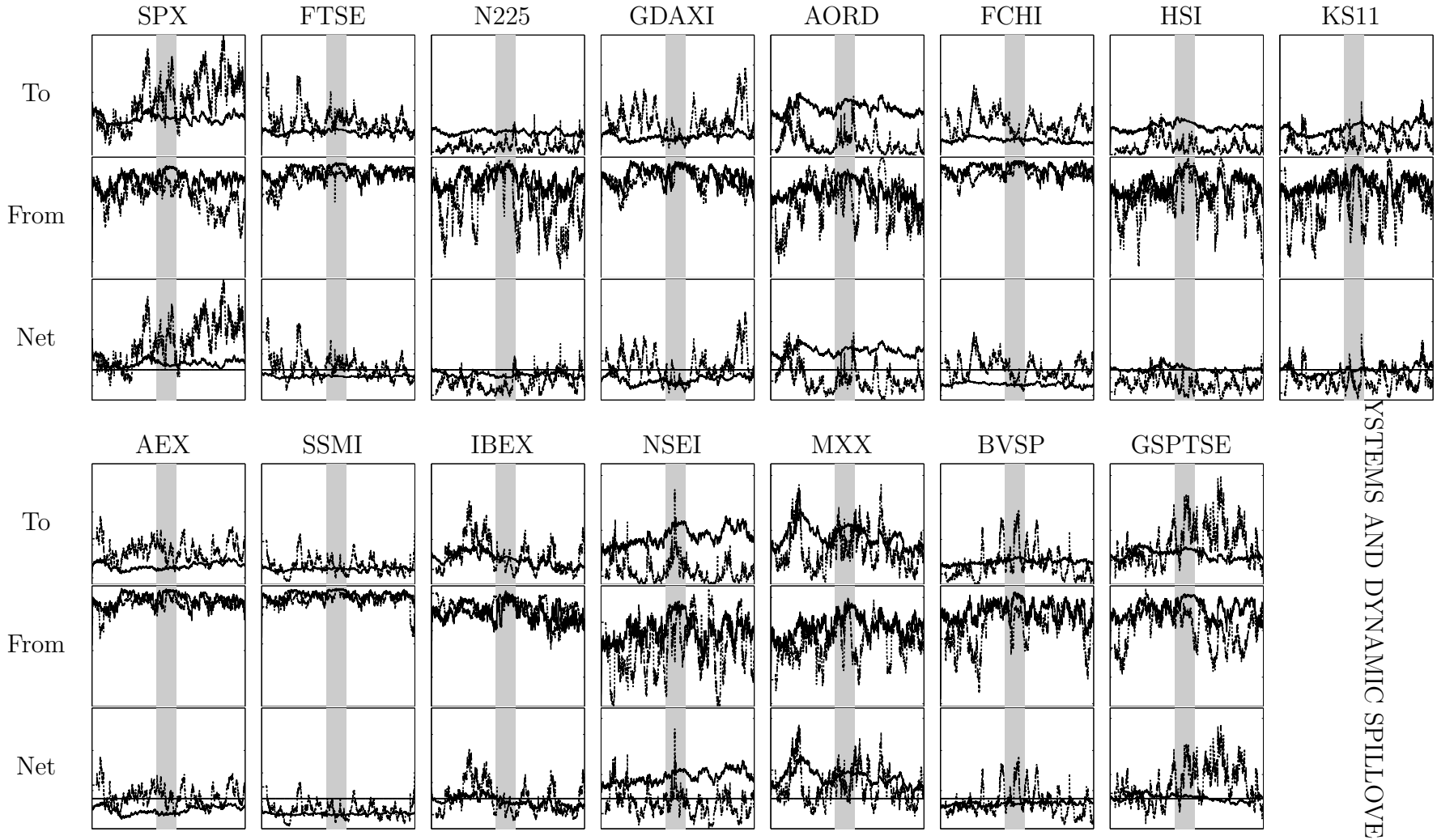


Figure 5.12: Individual volatility spillover series

Notes: The solid lines correspond to the TvpVAR model, while the dashed lines correspond to a rolling window of 100 days.

5.C Handling missing observations

We collect below the necessary notation and implementation details for a time-varying parameter VAR in an environment where the number of missing observations can follow any arbitrary patterns. In a n-dimensional vector autoregression of the possibly only partially observed data vector y_t, the number of available observations at time t is denoted by n_t. Furthermore, define the $(n_t \times n)$-dimensional selection matrix S_t with the property that $S_t y_t$ collects the non-missing observations, while $(S_t^\top S_t)^\perp$ is spanned by missing observations. The available data can then be modeled by

$$S_t y_t = \left(S_t \Phi_t S_{t-1}^\top\right) S_{t-1} y_{t-1} + S_t u_t,$$

or alternatively, in terms of a dynamic regression

$$S_t y_t = \left(\mathrm{I}_{n_t} \otimes \left(S_{t-1} y_{t-1}\right)^\top\right) \times \left(S_t \otimes S_{t-1}\right) \mathrm{vec}\left(\Phi_t^\top\right) + S_t u_t,$$

such that we adopt non-zero innovations for coefficients selected by $S_t \otimes S_{t-1}$.

Chapter 6

Conclusion

While the introduction had put the individual chapters into a common perspective, we wrap up by first specifically highlighting some of the most central results of this monograph. Finally, we critically appraise the content of this thesis by discussing extensions that could be considered for future research.

6.1 Summary of contributions

In the second chapter we have studied score driven correlation models from the viewpoint of stochastic recurrence equations. As the conditions for stochastic fixed points seem both analytically and numerically intangible, we need to restrict the parameter space in favor of making the conditions manageable. The simplification has some undesirable side-effects as it has introduced an indeterminacy regarding square

roots of the stochastic correlation matrix, which is potentially cumbersome as the indeterminacy is with respect to functions that index the relevant matrix decompositions. Fortunately, it turns out that across all possible Cholesky factorizations, only the symmetric matrix root has the property of cancelling out the undetermined terms in the stochastic fixed point conditions. Furthermore, we develop an efficient algorithm of verifying the new stochastic fixed point conditions and empirically illustrate the workings of score driven correlation models. It turns out that the indeterminacy regarding the appropriate Cholesky root can play a decisive role regarding the empirical usefulness of our simplified approach.

We proceed in Chapter 3 by developing an alternative approach that would be applicable for analyzing high-dimensional dynamic correlation models. On the one hand, the proposed approach delivers a precise and an intuitive description of the stability properties of a wide class of score driven models. On the other hand, while Chapter 2 had the benefit of guaranteeing pathwise uniqueness of the stable solution, which also nicely delivers invertibility conditions and maximum likelihood estimation, our statements in terms of geometric ergodicity are of distributional nature and need to condition upon invertibility of the model. We have succeeded in establishing some theoretical results regarding geometric ergodicity and a mis-specification robust quasi maximum likelihood estimation routine. As an empirical modeling strategy, we follow a radically simplistic correlation modeling strategy, driven by a single state variable. We show how our approach can improve upon both score driven and non-score driven

methods in the literature.

The fourth chapter extends the class of score driven models to allow for changes in deep, statistically structural, parameters of the model. We recognize that the GAS filter is a particularly well suited, but yet still mis-specified method of capturing stylized facts of various economic data. As the observation driven approach reveals the latent state variables as observable quantities, we can directly measure the empirical counterparts of the model-implied moment conditions on the state variables. This enables us to adjust the model transition dynamics to account for conditional and unconditional filter misspecifications. We demonstrate how the new filtering method improves upon the forecasting performance of highly competitive GAS models across different economic applications.

The fifth chapter deals with monitoring price spillover externalities in equity markets. We show how proper models for dynamic predictive systems can deliver superior forecasts and analyze the robustness of the spillover measures to the choice of filtering models. If a better description of the predictive relations translates to better systemic risk measures, then our proposed approach may deliver an attractive tool for policy makers.

6.2 Further extensions

A common theme of this thesis has been to understand specific models for multivariate financial time series. Even within this narrow field, our contributions to the literature have clear limitations, which we will highlight next. Such a discussion also clarifies the potential for further research.

First of all, understanding high-dimensional dynamic correlation models still poses a challenge to the econometrics profession. While diagnostic tests for residual cross-correlation undeniably reject our strict equicorrelation restriction in Chapter 3, it is still unclear in which direction the simplistic model should be generalized. The desired alternative model should deliver interpretable estimates, yet the practical estimation considerations require such a model to be sufficiently parsimonious. An interesting extension would be to allow for multiple correlation factors, where parsimony would be enforced within suitably chosen clusters. For instance, one could cluster individual equity return correlations across industry groups. However, the industry classification may not always deliver good descriptions – think for example of General Electric, which recently has been described as more of a financial conglomerate than an industrial company. Furthermore, it is unclear how to reconcile easily interpretable clustering methods with positive definiteness requirements of correlation matrices.

Second, the difficulties we have faced in justifying frequentist estimation methods beg the question whether Bayesian methods could prove useful in the context of score driven models. We have focused on consistency arguments, while asymptotic

distributional properties may have small-sample distortions in the face of heavy tailed observations and the often observed near unit root behavior of time-varying parameters. Furthermore, it would be interesting to further investigate whether score driven models could benefit from the usual merits of a Bayesian approach that may possibly guard one against overfitting in the face of parameter uncertainty.

Third, it would be worthwhile to give a second thought to the concept of statistical efficiency in high-dimensional parsimonious models. In Chapter 3, we considered Gaussian quasi maximum likelihood estimation of the static parameters in order to remain guarded against mis-specification of the observation density. The Gaussianity assumption is undeniably not the best description of the data, however due to the large amounts of data we would not expect the choice of the statistical criterion function to be a first order issue.

Fourth, it would be a worthy challenge to develop asymptotic theory for the tvGAS models in Chapter 4. A further interesting extension would be to consider more concurrent sources of misspecification.

Finally, the interpretation of spillovers in dynamic predictive systems seems nebulous: it is not entirely clear how to give the spillovers an intuitive meaning. The spillover centrality of a single asset could be driven by its high loading on a particular underlying risk factor. For this reason, an appealing alternative approach would be to consider time-varying factor augmented vector autoregressive (FAVAR) models to uncover the spillover dynamics. Such an approach would require one to work out the

spillover interpretations of the FAVAR-based variance decomposition. In particular, if the factors are spanned by fundamental financial and macroeconomic quantities, then the increased magnitudes of spillovers in times of financial crises could be given the true meaning of contagion.

Bibliography

Abadir, K. M. and J. R. Magnus (2005). *Matrix algebra*, Volume 1. Cambridge Univ Press.

Ahern, K. (2014). Network centrality and the cross section of stock returns. *Available at SSRN 2197370*.

Aielli, G. P. (2013). Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics 31*(3), 282–299.

Aït-Sahalia, Y. and T. R. Hurd (2014). Portfolio choice in markets with contagion. *Available at SSRN 2157038*.

Amado, C. and T. Teräsvirta (2013). Modelling volatility by variance decomposition. *Journal of Econometrics 175*(2), 142–153.

Andersen, T. G., T. Bollerslev, P. Christoffersen, and F. X. Diebold (2007). Practical volatility and correlation modeling for financial market risk management. In *The risks of financial institutions*, pp. 513–548. University of Chicago Press.

BIBLIOGRAPHY

Avdulaj, K. and J. Barunik (2015). Are benefits from oil-stocks diversification gone? New evidence from a dynamic copula and high frequency data. *Working paper, Charles University Prague*.

Barndorff-Nielsen, O. E. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(2), 253–280.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 1481–1536.

Bauwens, L., C. M. Hafner, and D. Pierret (2013). Multivariate volatility modeling of electricity futures. *Journal of Applied Econometrics 28*(5).

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics 21*(1), 79–109.

Bec, F., A. C. Rahbek, and N. Shephard (2008). The ACR model : A multivariate dynamic mixture autoregression. *Oxford Bulletin of Economics and Statistics 70*(5), 583–606.

Bekaert, G., M. Ehrmann, M. Fratzscher, and A. Mehl (2014). The global crisis and equity market contagion. *The Journal of Finance 69*(6), 2597–2649.

BIBLIOGRAPHY

Bernardi, M. and L. Catania (2015). Switching-GAS copula models for systemic risk assessment. *Working paper, University of Padova*.

Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society 12*(5), 788–792.

Billingsley, P. (1995). *Probability and measure*. Wiley series in probability and mathematical statistics.

Billio, M., M. Getmansky, D. Gray, A. Lo, R. C. Merton, and L. Pelizzon (2012). Sovereign, bank and insurance credit spreads: Connectedness and system networks. In *G-20 Conference on Financial Systemic Risk*, Volume 27.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics 104*(3), 535–559.

Bisias, D., M. D. Flood, A. W. Lo, and S. Valavanis (2012). A survey of systemic risk analytics. *US Department of Treasury, Office of Financial Research* (0001).

Black, F. (1976). Studies of stock, price volatility changes. *Proceedings of the 1976 meetings of the Business and Economics Statistics Association*, 177–181.

Blasques, F., S. J. Koopman, and A. Lucas (2014a). Maximum likelihood estimation for Generalized Autoregressive Score models. *Tinbergen Institute Discussion Papers 14-029/III*.

BIBLIOGRAPHY

Blasques, F., S. J. Koopman, and A. Lucas (2014b). Stationarity and ergodicity of univariate Generalized Autoregressive Score processes. *Electronic Journal of Statistics 8*, 1088–1112.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika*, forthcoming.

Blasques, F., A. Lucas, and E. Silde (2016). A stochastic recurrence equations approach for score driven correlation models. *Econometric Reviews*, forthcoming.

Boisbunon, A., S. Canu, D. Fourdrinier, W. Strawderman, and M. Wells (2014). AIC, Cp and estimators of loss for elliptically symmetric distributions. *International Statistical Review*. in press.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics 72*, 498–505.

Boudt, K., C. Croux, and S. Laurent (2011). Outlyingness weighted covariation. *Journal of Financial Econometrics 9*(4), 657–684.

Boudt, K., J. Danielsson, S. J. Koopman, and A. Lucas (2012). Regime switches in the volatility and correlation of financial institutions. Technical report, National Bank of Belgium.

BIBLIOGRAPHY

Boudt, K., J. Danielsson, and S. Laurent (2013). Robust forecasting of dynamic conditional correlation GARCH models. *International Journal of Forecasting 29*(2), 244–257.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Boussama, F. (1998). *Ergodicity, mixing and estimation in GARCH models*. Ph. D. thesis, University of Paris 7.

Boussama, F., F. Fuchs, and R. Stelzer (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications 121*(10), 2331–2360.

Buraschi, A., P. Porchia, and F. Trojani (2010). Correlation risk and optimal portfolio choice. *The Journal of Finance 65*(1), 393–420.

Buss, A. and G. Vilkov (2012). Measuring equity risk with option-implied correlations. *Review of Financial Studies 25*(10), 3113–3140.

Callot, L. A. and A. B. Kock (2014). Oracle efficient estimation and forecasting with the adaptive LASSO and the adaptive group LASSO in vector autoregressions. *Essays in Nonlinear Time Series Econometrics*, 238.

Calvet, L. E., V. Czellar, and E. Ronchetti (2014). Robust filtering. *Journal of the American Statistical Association* (in press).

BIBLIOGRAPHY

Chamberlain, G. (1983). A characterization of the distributions that imply mean–variance utility functions. *Journal of Economic Theory 29*(1), 185–201.

Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics 26*(6), 922–947.

Christoffersen, P., V. Errunza, K. Jacobs, and H. Langlois (2012). Is the potential for international diversification disappearing? A dynamic copula approach. *Review of Financial Studies 25*(12), 3711–3751.

Christoffersen, P., K. Jacobs, and B. Y. Chang (2011). Forecasting with option implied information. *CREATES Research Papers 46*.

Cline, D. and H. Pu (1999). Geometric ergodicity of nonlinear time series. *Statistica Sinica 9*(4), 1103–1118.

Cogley, T. and T. J. Sargent (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic dynamics 8*(2), 262–302.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-

tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics 29*(4), 552–563.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized Autoregressive Score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., S. J. Koopman, A. Lucas, and B. Schwaab (2014). Observation-driven mixed measurement dynamic factor models. *Review of Economics and Statistics*, forthcoming.

Creal, D. D. and R. S. Tsay (2014). High dimensional dynamic stochastic copula models. Technical report, Chicago Booth Working Paper.

Dahlhaus, R. (2012). Locally stationary processes. *Handbook of Statistics, Time Series Analysis: Methods and Applications*, 351–408.

Dahlhaus, R. et al. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics 25*(1), 1–37.

Dahlhaus, R., S. S. Rao, et al. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics 34*(3), 1075–1114.

Davidson, R. and J. G. MacKinnon (2004). *Econometric theory and methods*, Volume 5. Oxford University Press New York.

De Lira Salvatierra, I. and A. J. Patton (2013). Dynamic copula models and high frequency data. *Duke University Working Paper*.

BIBLIOGRAPHY

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

DeMiguel, V., Y. Plyakha, R. Uppal, and G. Vilkov (2013). Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis 48*(06), 1813–1845.

Diaconis, P. and D. Freedman (1999). Iterated random functions. *SIAM review*, 45–76.

Diebold, F. X. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Reviews 5*(1), 51–56.

Diebold, F. X. and G. D. Rudebusch (1994). Measuring business cycles: A modern perspective. Technical report, National Bureau of Economic Research.

Diebold, F. X. and K. Yilmaz (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal 119*(534), 158–171.

Diebold, F. X. and K. Yılmaz (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*.

Ding, Z. (1994). *Time Series Analysis of Speculative Returns*. PhD dissertation, University of California, San Diego.

BIBLIOGRAPHY

Driessen, J., P. J. Maenhout, and G. Vilkov (2009). The price of correlation risk: Evidence from equity options. *The Journal of Finance 64*(3), 1377–1406.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Number 38. Oxford University Press.

Engle, R. (2002a). Dynamic Conditional Correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. (2002b). New frontiers for ARCH models. *Journal of Applied Econometrics 17*(5), 425–446.

Engle, R. and K. Kroner (1995a). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*, 122–150.

Engle, R. and G. Lee (1999). A permanent and transitory component model of stock return volatility. *in ed. R.F. Engle and H. White, Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, 475–497.

Engle, R., N. Shephard, and K. Sheppard (2008). Fitting vast dimensional time-varying covariance models.

Engle, R. F. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business and Economic Statistics 30*, 212–228.

Engle, R. F. and K. F. Kroner (1995b). Multivariate simultaneous generalized ARCH. *Econometric Theory 11*(01), 122–150.

Engle, R. F. and V. K. Ng (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance 48*(5), 1749–1778.

Forbes, K. J. and R. Rigobon (2002). No contagion, only interdependence: measuring stock market comovements. *The Journal of Finance 57*(5), 2223–2261.

Francq, C. and J. Zakoian (2011). *GARCH models: structure, statistical inference and financial applications*. Wiley.

Francq, C. and J.-M. Zakoïan (2012). Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica 80*(2), 821–861.

Gallant, A. R. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*(04), 657–681.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Giglio, S., B. T. Kelly, and S. Pruitt (2015). Systemic risk and the macroeconomy: An empirical evaluation. Technical report, National Bureau of Economic Research.

Giraitis, L., G. Kapetanios, K. Theodoridis, and T. Yates (2014). Estimating time-varying DSGE models using minimum distance methods.

BIBLIOGRAPHY

Giraitis, L., G. Kapetanios, and T. Yates (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics 179*(1), 46–65.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance 48*(5), 1779–1801.

Goodwin, G. C. and K. S. Sin (2014). *Adaptive filtering prediction and control*. Courier Corporation.

Granger, C. W. (2001). Forecasting white noise. *Econometric Society Monographs 32*, 457–471.

Granger, C. W. (2008). Non-linear models: Where do we go next-time varying parameter models? *Studies in Nonlinear Dynamics & Econometrics 12*(3).

Haas, M., S. Mittnik, and M. S. Paolella (2004). Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics 2*(2), 211–250.

Hamada, M. and E. A. Valdez (2008). CAPM and option pricing with elliptically contoured distributions. *Journal of Risk and Insurance 75*(2), 387–409.

Hansen, B. E. (1992, August). Testing for parameter instability in linear models. *Journal of Policy Modeling 14*(4), 517–533.

Hansen, P. R., Z. Huang, and H. H. Shek (2012). Realized GARCH: a joint model for

returns and realized measures of volatility. *Journal of Applied Econometrics 27*(6), 877–906.

Hansen, P. R. and A. Lunde (2014). Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory 30*(01), 60–93.

Hardle, W. (1990). *Applied nonparametric regression*, Volume 27. Cambridge Univ Press.

Harvey, A. and T. Chakravarty (2008). Beta-t-(e)GARCH. *University of Cambridge, Faculty of Economics, Working paper CWPE 08340*.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press.

Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, forthcoming.

Hautsch, N., J. Schaumburg, and M. Schienle (2014). Financial network systemic risk contributions. *Review of Finance*, in press.

Heber, G., A. Lunde, N. Shephard, and K. Sheppard (2009). OMIs realised library, version 0.1.

Janus, P., S. J. Koopan, and A. Lucas (2014). Long memory dynamics for multivariate dependence under heavy tails. *Journal of Empirical Finance*, forthcoming.

Jensen, A. T. and T. Lange (2010). On convergence of the QMLE for misspecified GARCH models. *Journal of Time Series Econometrics 2*(1).

Kleibergen, F. and H. K. van Dijk (1993). Non-stationarity in garch models: A bayesian analysis. *Journal of Applied Econometrics 8*(S1), S41–S61.

Klößner, S. and S. Wagner (2014). Exploring all VAR orderings for calculating spillovers? yes, we can!A note on Diebold and Yilmaz (2009). *Journal of Applied Econometrics 29*(1), 172–179.

Klüppelberg, C., A. Lindner, and R. Maller (2004). A continuous-time GARCH process driven by a lévy process: Stationarity and second-order behaviour. *Journal of Applied Probability 41*(3), 601–622.

Kodres, L. E. and M. Pritsker (2002). A rational expectations model of financial contagion. *The Journal of Finance 57*(2), 769–799.

Koijen, R. S. J., T. J. Moskowitz, L. H. Pedersen, and E. B. Vrugt (2015). Carry. Technical report, NBER Working Paper Series.

Koo, B. and O. Linton (2012). Estimation of semiparametric locally stationary diffusion models. *Journal of Econometrics 170*(1), 210–233.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics 74*(1), 119–147.

BIBLIOGRAPHY

Koopman, S., F. Blasques, and Z. Zhang (2015). Low frequency and weighted likelihood solutions for mixed frequency dynamic factor models. Technical report, Tinbergen Institute Working Paper Series.

Koopman, S. J., A. Lucas, and M. Scharth (2012). Predicting time-varying parameters with parameter-driven and observation-driven models. *Tinbergen Institute Discussion Papers 12-020/4*.

Krengel, U. (1985). *Ergodic theorems*. Berlin: De Gruyter studies in Mathematics.

Laurent, S., J. V. Rombouts, and F. Violante (2009). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*.

Liu, L. Y., A. J. Patton, and K. Sheppard (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics 187*(1), 293–311.

Lucas, A., B. Schwaab, and X. Zhang (2013). Measuring credit risk in a large banking system: Econometric modeling and empirics. Technical report, Tinbergen Institute Discussion Paper 13-063/IV/DSF56.

Lucas, A., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics 32*(2), 271–284.

BIBLIOGRAPHY

Nelson, D. B. (1990). Stationarity and persistence in the GARCH (1, 1) model. *Econometric Theory 6*(03), 318–334.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

Oh, D. H. and A. J. Patton (2012). Modelling dependence in high dimensions with factor copulas. *Manuscript, Duke University*.

Oh, D. H. and A. J. Patton (2013). Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads.

Owen, J. and R. Rabinovitch (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance 38*(3), 745–752.

Patton, A. J. (2006a). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics 21*(2), 147–173.

Patton, A. J. (2006b). Modelling asymmetric exchange rate dependence. *International Economic Review 47*(2), 527–556.

Patton, A. J. (2009). Copula–based models for financial time series. *Handbook of Financial Time Series*, 767–785.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics 160*(1), 246–256.

Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics 131*(1), 445–473.

Pesaran, H. and Y. Shin (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters 58*(1), 17 – 29.

Pesaran, M. H., A. Pick, and M. Pranovich (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics 177*(2), 134–152.

Pollet, J. M. and M. Wilson (2010). Average correlation and stock market returns. *Journal of Financial Economics 96*(3), 364–380.

Rahbek, A. and R. S. Pedersen (2013). Multivariate variance targeting in the BEKK-GARCH model. *Econometrics Journal*.

Shephard, N., D. Noureldin, and K. K. Sheppard (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 27*, 907–933.

Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH models. In *Handbook of Financial Time Series*, pp. 201–229. Springer.

Sims, C. (2001). Comment on Sargent and Cogleys: Evolving Post World War II U.S. Inflation Dynamics. *NBER Macroeconomics Annual*, 373–379.

Stărică, C. and C. Granger (2005). Nonstationarities in stock returns. *Review of Economics and Statistics 87*(3), 503–522.

BIBLIOGRAPHY

Stock, J. H. and M. W. Watson (2005). Understanding changes in international business cycle dynamics. *Journal of the European Economic Association 3*(5), 968–1006.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroeskedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, 1–18.

Van der Vaart, A. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics 17*(5), 549–564.

van Dijk, D., S. J. Koopman, M. Wel, and J. H. Wright (2014). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics*.

Vasnev, A. L., G. Claeskens, J. Magnus, and W. Wang (2014). A simple theoretical explanation of the forecast combination puzzle. *Tinbergen Institute Discussion paper 14-127/III*.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

BIBLIOGRAPHY

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge Books. Cambridge University Press.

Wu, W. and X. Shao (2004). Limit theorems for iterated random functions. *Journal of Applied Probability 41*(2), 425–436.

Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

Summary

This dissertations presents four articles in four chapters. Each chapter adopts various statistical time series methods to analyze co-movement in financial markets. In particular, the following four questions are answered: 1) What are the statistical properties of GAS (Generalized Autoregressive Score) time series correlation models? 2) How can we model correlations in high-dimensional datasets? 3) How can we adapt dynamic filtering models, if we see that the model does not resemble characteristics of the data? 4) If we monitor financial spillovers, how can we dynamically model its characteristics?

Chapter 2 applies a simulation-based method to answer the first question. We discover that the non-uniquess of matrix square root introduces an additional ingredient to determine the stability of the model. Moreover, we find that the symmetric matrix square root offers the best overall results.

Chapter 3 introduces a model for high-dimensional dynamic correlation data. This model can be estimated by quasi-Gaussian maximum-likelihood methods, which ensures the usual statistical optimality properties. We demonstrate numerical efficiency in evaluation of the model. We also develop a dynamic correlation model which incorporates data from both option markets and equity markets. This enables us to monitor correlation risk premia in real time.

Chapter 4 presents a systematic approach to adopt filtering equations to misspecification of the model. Once we recognize the misfit of the dynamic filtering model with respect to a concrete property of the data, we are able to come up with a new, improved model. We demonstrate the usefulness of our method using various practical examples.

Chapter 5 introduces a dynamic vector autoregressive model to monitor spillover effects in financial markets. So far researchers have used ad hoc rolling windows methods to analyze comovement connectedness networks. The new model provides more accurate predictions and yields different, more reactive, measures of spillovers.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

641 T. HOMAR, *Intervention in Systemic Banking Crises*

642 R. LIT, *Time Varying Parameter Models for Discrete Valued Time Series*

643 R.H. KLEIJN, *Essays on Bayesian Model Averaging using Economic Time Series*

644 S. MUNS, *Essays on Systemic Risk*

645 B.M. SADABA, *Essays on the Empirics of International Financial Markets*

646 H. KOC, *Essays on Preventive Care and Health Behaviors*

647 V.V.M. MISHEVA, *The Long Run Effects of a Bad Start*

648 W. LI, *Essays on Empirical Monetary Policy*

649 J.P. HUANG, *Topics on Social and Economic Networks*

650 K.A. RYSZKA, *Resource Extraction and the Green Paradox: Accounting for Political Economy Issues and Climate Policies in a Heterogeneous World*

651 J.R. ZWEERINK, *Retirement Decisions, Job Loss and Mortality*

652 M. K. KAGAN, *Issues in Climate Change Economics: Uncertainty, Renewable Energy Innovation and Fossil Fuel Scarcity*

653 T.V. WANG, *The Rich Domain of Decision Making Explored: The Non-Triviality of the Choosing Process*

654 D.A.R. BONAM, *The Curse of Sovereign Debt and Implications for Fiscal Policy*

655 Z. SHARIF, *Essays on Strategic Communication*

656 B. RAVESTEIJN, *Measuring the Impact of Public Policies on Socioeconomic Disparities in Health*

657 M. KOUDSTAAL, *Common Wisdom versus Facts; How Entrepreneurs Differ in Their Behavioral Traits from Others*

658 N. PETER, *Essays in Empirical Microeconomics*

659 Z. WANG, *People on the Move: Barriers of Culture, Networks, and Language*

660 Z. HUANG, *Decision Making under Uncertainty-An Investigation from Economic and Psychological Perspective*

661 J. CIZEL, *Essays in Credit Risk, Banking, and Financial Regulation*

662 I. MIKOLAJUN, *Empirical Essays in International Economics*

663 J. BAKENS, *Economic Impacts of Immigrants and Ethnic Diversity on Cities*

664 I. BARRA, *Bayesian Analysis of Latent Variable Models in Finance*

665 S. OZTURK, *Price Discovery and Liquidity in the High Frequency World*

666 J. JI, *Three Essays in Empirical Finance*

667 H. SCHMITTDIEL, *Paid to Quit, Cheat, and Confess*

668 A. DIMITROPOULOS, *Low Emission Vehicles: Consumer Demand and Fiscal Policy*

669 G.H. VAN HEUVELEN, *Export Prices, Trade Dynamics and Economic Development*

670 A. RUSECKAITE, *New Flexible Models and Design Construction Algorithms for Mixtures and Binary Dependent Variables*

671 Y. LIU, *Time-varying Correlation and Common Structures in Volatility*

672 S. HE, *Cooperation, Coordination and Competition: Theory and Experiment*

673 C.G.F. VAN DER KWAAK, *The Macroeconomics of Banking*

674 D.H.J. CHEN, *Essays on Collective Funded Pension Schemes*

675 F.J.T. SNIEKERS, *On the Functioning of Markets with Frictions*

676 F. GOMEZ MARTINEZ, *Essays in Experimental Industrial Organization: How Information and Communication affect Market Outcomes*

677 J.A. ATTEY, *Causes and Macroeconomic Consequences of Time Variations in Wage Indexation*

678 T. BOOT, *Macroeconomic Forecasting under Regime Switching, Structural Breaks and High-dimensional Data*

679 I. TIKOUDIS, *Urban Second-best Road Pricing: Spatial General Equilibrium Perspectives*

680 F.A. FELS, *Empirical Studies of Consumer and Government Purchase Decisions*

681 Y. GAO, *Stability and Adaptivity: Preferences over Time and under Risk*

682 M.J. ZAMOJSKI, *Panta Rhei, Measurement and Discovery of Change in Financial Markets*

683 P.R. DENDERSKI, *Essays on Information and Heterogeneity in Macroeconomics*

684 U. TURMUNKH, *Ambiguity in Social Dilemmas*

685 U. KESKIN, *Essays on Decision Making: Intertemporal Choice and Uncertainty*

686 M. LAMMERS, *Financial Incentives and Job Choice*

687 Z. ZHANG, *Topics in Forecasting Macroeconomic Time Series*

688 X. XIAO, *Options and Higher Order Risk Premiums*

689 D.C. SMERDON, *Everybodys doing it: Essays on Trust, Norms and Integration*

690 S. SINGH, *Three Essays on the Insurance of Income Risk and Monetary Policy*